

Stable states of Perturbed Markov Chains

Volker Betz

Technische Universität Darmstadt
betz@mathematik.tu-darmstadt.de

Stéphane Le Roux *

Université libre de Bruxelles
stephane.le.roux@ulb.ac.be

Abstract

Given an infinitesimal perturbation of a discrete-time finite Markov chain, we seek the states that are stable despite the perturbation, *i.e.* the states whose weights in the stationary distributions can be bounded away from 0 as the noise fades away. Chemists, economists, and computer scientists have been studying irreducible perturbations built with exponential maps. Under these assumptions, Young proved the existence of and computed the stable states in cubic time. We fully drop these assumptions, generalize Young's technique, and show that stability is decidable as long as $f \in O(g)$ is. Furthermore, if the perturbation maps (and their multiplications) satisfy $f \in O(g)$ or $g \in O(f)$, we prove the existence of and compute the stable states and the metastable dynamics at all time scales where some states vanish. Conversely, if the big- O assumption does not hold, we build a perturbation with these maps and no stable state. Our algorithm also runs in cubic time despite the general assumptions and the additional work. Proving the correctness of the algorithm relies on new or rephrased results in Markov chain theory, and on algebraic abstractions thereof.

Categories and Subject Descriptors G [3]

General Terms

Keywords evolution, learning, metastability, tropical algebra, shortest path, SCC, cubic time algorithm

1. Introduction

Motivated by the dynamics of chemical reactions, Eyring [4] and Kramers [12] studied how infinitesimal perturbations of a Markov chain affects its stationary distributions. This topic has been further investigated by several academic communities including probability theorists, economists, and computer scientists. In several fields of application, such as learning and game theory, it is sometimes unnecessary to describe the exact values of the limit stationary distributions: it suffices to know whether these values are zero or not. Thus, the *stochastically stable states* ([5], [10], [19]) were defined in different contexts as the states that have positive probability in the limit. We rephrase a definition below.

* This author is supported by the ERC inVEST (279499) project and was in TU Darmstadt when this research started.

Definition 1 (Markov chain perturbation and stochastic stability)

Let I be a subset of positive real numbers with 0 as a limit point for the usual topology¹. A perturbation is a family $((X_n^{(\epsilon)})_{n \in \mathbb{N}})_{\epsilon \in I}$ of discrete-time Markov chains sharing the same finite state space. If the chain $(X_n^{(\epsilon)})_{n \in \mathbb{N}}$ is irreducible for all $\epsilon \in I$, then $((X_n^{(\epsilon)})_{n \in \mathbb{N}})_{\epsilon \in I}$ is said to be an irreducible perturbation.

A state x of $((X_n^{(\epsilon)})_{n \in \mathbb{N}})_{\epsilon \in I}$ is stochastically stable if there exists a family of corresponding stationary distributions $(\mu_\epsilon)_{\epsilon \in I}$ such that $\liminf_{\epsilon \rightarrow 0} \mu_\epsilon(x) > 0$. It is stochastically fully vanishing if $\limsup_{\epsilon \rightarrow 0} \mu_\epsilon(x) = 0$ for all $(\mu_\epsilon)_{\epsilon \in I}$. Non-stable states are called vanishing.

Definition 1 may be motivated in at least two ways. First, a dynamical system (*e.g.* modeled by a Markov chain) has been perturbed from the outside, and the laws governing the systems (*e.g.* the transition probability matrix) have been changed. As time elapses (*i.e.* as ϵ approaches zero), the laws slowly go back to normal. What are the almost sure states of the system after infinite time? Second, a very complex Markov chain is the sum of a simple chain and a complex perturbation matrix that is described *via* a small, fixed ϵ_0 . The stationary distributions of the complex chain are hard to compute, but which states have significantly positive probability after infinite time? Our main result below answers these questions.

Theorem 2 Consider a perturbation such that $f \in O(g)$ or $g \in O(f)$ for all f and g in the multiplicative closure of the transition probability functions $\epsilon \mapsto p_\epsilon(x, y)$ with $x \neq y$. Then the perturbation has stable states, and stability can be decided in $O(n^3)$, where n is the number of states.

Note that by finiteness of the state space it is easy to prove that every perturbation has a state that is not fully vanishing.

1.1 Related works and comparisons

In 1990 Foster and Young [5] defined the stochastically stable states of a general (continuous) evolutionary process, as an alternative to the evolutionary stable strategies [16]. Stochastically stable states were soon adapted by Kandori, Mailath, and Rob [10] for evolutionary game theory with 2×2 games. Then Young [19, Theorem 4] proved "a finite version of results obtained by Freidlin and Wentzel" in [6]. Namely, he characterized the stochastically stable states if the perturbation satisfies the following assumptions: 1) the perturbed matrices P^ϵ are aperiodic and irreducible; 2) the P^ϵ converge to the unperturbed matrix P^0 when ϵ approaches zero; 3) every transition probability is a function of ϵ that is equivalent to $c \cdot \epsilon^\alpha$ for some non-negative real numbers c and α . The main tool in Young's proof was proved by Kohler and Vollmerhaus [11] and is the special case for irreducible chains of the Markov chain tree theorem (see [13] or [6]). Young's characterization involves minimum directed spanning trees, which can be computed in $O(n^2)$ [7] for

¹ This implies that I is infinite. $]0, 1]$ and $\{\frac{1}{2^n} \mid n \in \mathbb{N}\}$ are typical I .

graphs with n vertices. Since there are at most n roots for directed spanning trees in a graph with n vertices, Young can compute the stable states in $O(n^3)$.

In 2000, Ellison [3] characterized the stable states *via* the alternative notion of the radius of a basin of attraction. The major drawback of his characterization compared to Young's is that it is "not universally applicable" [3]; the advantages are that it provides "a bound on the convergence rate as well as a long-run limit" and "intuition for why the long-run stochastically stable set of a model is what it is". In 2005, Wicks and Greenwald [9] designed an algorithm to express the exact values of the limit stationary distribution of a perturbation, which, as a byproduct, also computes the set of the stable states. Like [19] they consider perturbations that are related to the functions $\epsilon \mapsto \epsilon^\alpha$, but they only require that the functions converge exponentially fast. Also, instead of requiring that the P^ϵ be irreducible for $\epsilon > 0$, they only require that they have exactly one essential class. They do not analyze the complexity of their algorithm but it might be polynomial time. We improve upon [19], [3], and [9] in several ways.

1. The perturbation maps in the literature relate to the maps $\epsilon \mapsto \epsilon^\alpha$. Their specific form and their continuity, especially at 0, are used in the existing proofs. Theorem 2 dramatically relaxes this assumption. Continuity, even at 0, is irrelevant, which allows for aggressive, *i.e.*, non-continuous perturbations. We show that our assumption is (almost) unavoidable.
2. The perturbations in the literature are irreducible (but [9] slightly weakened this assumption). It is general enough for perturbations relating to the maps $\epsilon \mapsto \epsilon^\alpha$, since it suffices to process each sink (aka bottom) irreducible component independently, and gather the results. Although this trick does not work for general perturbation maps, Theorem 2 manages not to assume irreducibility.
3. The perturbation is abstracted into a weighted graph and shrunk by combining recursively a shortest-path algorithm (w.r.t. some tropical-like semiring) and a strongly-connected-component algorithm. Using tropical-like algebra to abstract over Markov chains has already been done before, but not to solve the stable state problem. ([8] did it to prove an algebraic version of the Markov chain tree theorem.)
4. Our algorithm computes the stable states in $O(n^3)$, as in [19], which is the best known complexity. In addition, the computation itself is a summary of the asymptotic behavior of the perturbation: it says at which time scales the vanishing states vanish, and the intermediate graph obtained at each recursive stage of the algorithm accounts for the metastable dynamics of the perturbation at this vanishing time scale.

Section 1.2 sets some notations; Section 1.3 analyses which assumptions are relevant for the existence of stable states; Section 2 proves the existential part of Theorem 2, *i.e.* it develops the probabilistic machinery to prove the existence of stable states; hinging on this, Section 3 proves the algorithmic part of Theorem 2, *i.e.* it abstracts the relevant objects using a new algebraic structure, presents the algorithm, and proves its correctness and complexity; Section 4 discusses two important special cases and an induction proof principle related to the termination of our algorithm.

1.2 Notations

- The set \mathbb{N} of the natural numbers contains 0. For a set S and $n \in \mathbb{N}$, let S^n be the words γ over S of length $|\gamma| = n$. Let $S^* := \cup_{n \in \mathbb{N}} S^n$ be the finite words over S . The set-theoretical notation $\cup E := \cup_{x \in E} x$ is used in some occasions.
- Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with state space S . For all $A \subseteq S$ let $\tau_A := \inf\{n \geq 0 : X_n \in A\}$ ($\tau_A^+ := \inf\{n > 0 : X_n \in A\}$) be the first time (first positive time) that the chain hits a state inside A . Usually $\tau_{\{x\}}$ and $\tau_{\{x\}}^+$ are written τ_x and τ_x^+ , respectively.

$X_n \in A\}$) be the first time (first positive time) that the chain hits a state inside A . Usually $\tau_{\{x\}}$ and $\tau_{\{x\}}^+$ are written τ_x and τ_x^+ , respectively.

- Given a Markov chain $(X_n)_{n \in \mathbb{N}}$, the corresponding matrix representation, law of the chain when started at state x , expectation when started at state x , and possible stationary distributions are respectively denoted p , \mathbb{P}^x , \mathbb{E}^x , and μ . When considering other Markov chains $(\tilde{X}_n)_{n \in \mathbb{N}}$ or $(\hat{X}_n)_{n \in \mathbb{N}}$, the derived notions are denoted with tilde or circumflex, as in \tilde{p} or $\hat{\mu}$.
- A perturbation $((X_n^{(\epsilon)})_{n \in \mathbb{N}})_{\epsilon \in I}$ will often be denoted X for short, and when it is clear from the context that we refer to a perturbation, p will denote the function $(\epsilon, x, y) \mapsto p_\epsilon(x, y)$ (instead of $(x, y) \mapsto p(x, y)$), and $p(x, y)$ will denote $\epsilon \mapsto p_\epsilon(x, y)$ (instead of a mere real number). The other derived notions are treated likewise.
- The probability of a path is defined inductively by $p(xy) := p(x, y)$ and $p(xy\gamma) := p(x, y)p(y\gamma)$ for all $x, y \in S$ and $\gamma \in S \times S^*$.
- Given x , y , and a set A , a simple A -path from x to y is a repetition-free (unless $x = y$) word γ starting with x and ending with y , and using beside x and y only elements in A . Formally, $\Gamma_A(x, y) := \{\gamma \in \{x\} \times A^* \times \{y\} \mid (1 \leq i < j \leq |\gamma| \wedge \gamma_i = \gamma_j) \Rightarrow (i = 1 \wedge j = |\gamma|)\}$.

1.3 Towards general assumptions

A state x of a perturbation is stable if there exists a related family $(\mu_\epsilon)_{\epsilon \in I}$ of stationary distributions such that $\mu(x) = O(1)$, but even continuous perturbations that converge when ϵ approaches 0 may fail to have stable states. For instance let $S := \{x, y\}$ and for all $\epsilon \in]0, 1]$ let $p_\epsilon(x, y) := \epsilon^2$ and $p_\epsilon(y, x) := \epsilon^{2+\cos(\epsilon^{-1})}$ as in Figure 1a, where the self-loops are omitted. In the unique stationary distribution x has a weight $\mu_\epsilon(x) = (1 + \epsilon^{-\cos(\epsilon^{-1})})^{-1}$. Since $\mu_{(2n\pi)^{-1}}(x) = \frac{2n\pi}{1+2n\pi} \rightarrow_{n \rightarrow \infty} 1$ and $\mu_{(2(n+1)\pi)^{-1}}(x) = \frac{1}{1+2(n+1)\pi} \rightarrow_{n \rightarrow \infty} 0$, neither x nor y is stable.

As mentioned above, the perturbations in the literature are related to the functions $\epsilon \mapsto \epsilon^\alpha$ with $\alpha \geq 0$, which rules out the example from Figure 1a and implies the existence of a stable state [19]. Here, however, we want to assume as little as possible about the perturbations, while still guaranteeing the existence of stable states. Towards it let us first rephrase the big O notation as a binary relation. It is well-known that big O enjoys various algebraic properties. The ones we need are mentioned in the appendix.

Definition 3 (Order) For $f, g : I \rightarrow [0, 1]$, let us write $f \precsim g$ if there exist positive b and ϵ such that $f(\epsilon') \leq b \cdot g(\epsilon')$ for all $\epsilon' < \epsilon$; let $f \cong g$ stand for $f \precsim g \wedge g \precsim f$.

Requiring that every two transition probability maps f and g occurring in the perturbation satisfy $f \precsim g$ or $g \precsim f$ rules out the example from Figure 1a, but not the one from Figure 1b. There $\mu_\epsilon(z) \leq \mu_\epsilon(x) = \frac{\epsilon^{\cos(\epsilon^{-1})}}{1 + \epsilon^{\cos(\epsilon^{-1})}(1 + \epsilon^2)}$ and $\mu_\epsilon(y) = \frac{1}{1 + \epsilon^{\cos(\epsilon^{-1})}(1 + \epsilon^2)}$. So $\mu_\epsilon(z) \rightarrow_{\epsilon \rightarrow 0} 0$ and $\mu_{2n\pi}(y) \rightarrow_{n \rightarrow \infty} 0$ and $\mu_{2(n+1)\pi}(x) \rightarrow_{n \rightarrow \infty} 0$, no state is stable. Informally, z is not stable because it gives everything but receives at most ϵ ; neither x nor y is stable since their interaction resembles Figure 1a due to ϵ^6 and $\epsilon^4 \cdot \epsilon^{2+\cos(\epsilon^{-1})}$. This remark is turned into a general Observation 4 below.

Observation 4 For $1 \leq i \leq n$ and $1 \leq j \leq m$ let $f_i, g_j : I \rightarrow [0, 1]$ be such that $\prod_i f_i$ and $\prod_j g_j$ are not \precsim -comparable. Then there exists a perturbation without stable states that is built only with the $f_1, \dots, f_n, g_1, \dots, g_m$ and the $1 - f_1, \dots, 1 - f_n, 1 - g_1, \dots, 1 - g_m$. See Figure 1d.

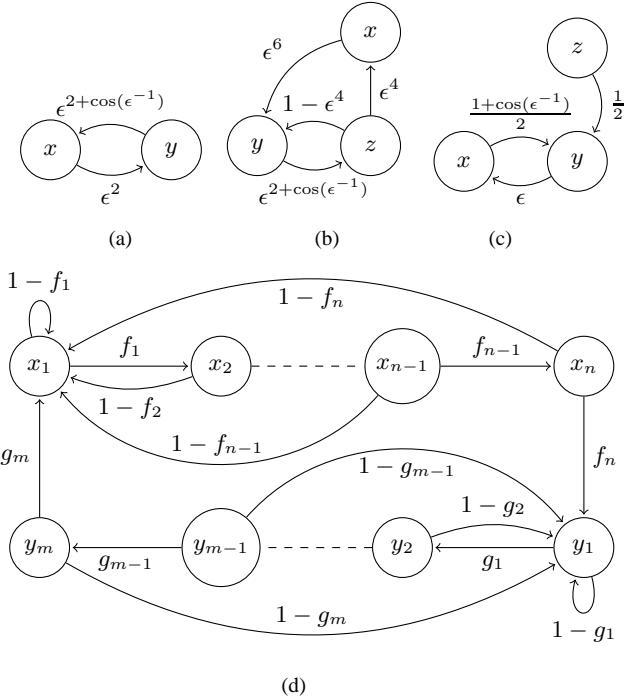


Figure 1: Perturbations without stable states

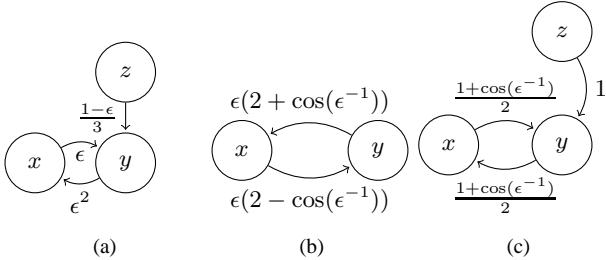


Figure 2: Perturbations with stable states

Observation 4 motivates the following "unavoidable" assumption.

Assumption 5 *The multiplicative closure of the maps $\epsilon \mapsto p_\epsilon(x, y)$ with $x \neq y$ is totally preordered by \precsim .*

For example, the classical maps $\epsilon \mapsto c \cdot \epsilon^\alpha$ with $c > 0$ and $\alpha \in \mathbb{R}$ constitute a multiplicative group totally preordered by \precsim . One reason why we can afford such a weak Assumption 5 is that we are not interested in the exact weights of some putative limit stationary distribution, but only whether the weights are bounded away from zero.

Let us show the significance of Assumption 5, which is satisfied by the perturbations in Figure 2 and 5e: Young's result shows that y is the unique stable state of the perturbation in Figure 2a, but it cannot say anything about Figures 2b, 2c, and 5e: Figure 2b is not regular, i.e., $\frac{2+\cos(\epsilon^{-1})}{2-\cos(\epsilon^{-1})}$ does not converge, and neither do the weights $\mu_\epsilon(x)$ and $\mu_\epsilon(y)$, but it is possible to show that both limits inferior are $1/4$ nonetheless, so both x and y are stable; the transition probabilities in Figure 2c do not converge, and $\frac{1+\cos(\epsilon^{-1})}{2}$ and $1 - \frac{1+\cos(\epsilon^{-1})}{2}$ are not even comparable, but it is easy to see that $\mu_\epsilon(x) = \mu_\epsilon(y) = \frac{1}{2}$; and in Figure 5e x is the only stable state

since its weight oscillates between $\frac{1}{2}$ and 1. Note that Assumption 5 rules out the perturbations in Figure 1, which have no stable state.

2. Existence of stable states

This section presents three transformations that simplify perturbations while retaining the relevant information about the stable states. Two of them are defined via the dynamics of the original perturbation. The relevance of these two transformations relies on the close relation between the stationary distributions and the dynamics of Markov chains. Lemma 6 below pinpoints this relation.

Lemma 6 *A distribution μ of a finite Markov chain is stationary iff its support involves only essential states and for all states x and y we have $\mu(x)\mathbb{P}^x(\tau_y^+ < \tau_x^+) = \mu(y)\mathbb{P}^y(\tau_x^+ < \tau_y^+)$.*

Lemma 6 can already help us find the stable states of small examples such as in Figures 1 and 2. In Figure 1a it says that $\mu_\epsilon(x)\epsilon^2 = \mu_\epsilon(y)\epsilon^{2+\cos(\epsilon^{-1})}$ so we find $\liminf \mu_\epsilon(x) = \liminf \mu_\epsilon(y) = 0$ without calculating the stationary distributions. In Figure 2b it says that $\mu_\epsilon(x)(2 - \cos(\epsilon^{-1})) = \mu_\epsilon(y)(2 + \cos(\epsilon^{-1}))$, so $\mu_\epsilon(x) \leq 3\mu_\epsilon(y)$ and $\frac{1}{4} \leq \mu_\epsilon(y)$, and likewise for x .

Lemma 7 below shows further connections between the stationary distributions and the dynamics of Markov chains. Its proof involves Lemma 6, and its irreducible case is used in Section 2.3.

Lemma 7 *Let p be a Markov chain with state space S , and let \tilde{p} be defined over $\tilde{S} \subseteq S$ by $\tilde{p}(x, y) := \mathbb{P}^x(X_{\tau_{\tilde{S}}^+} = y)$.*

1. Then $\mathbb{P}^x(\tau_y < \tau_x^+) = \tilde{\mathbb{P}}^x(\tau_y < \tau_x^+)$ for all $x, y \in \tilde{S}$.
2. Let $\mu(\tilde{\mu})$ be a stationary distribution for $p(\tilde{p})$. If \tilde{S} are essential states, there exists $\tilde{\mu}(\mu)$ a stationary distribution for $p(p)$ such that $\mu(x) = \tilde{\mu}(x) \cdot \sum_{y \in \tilde{S}} \mu(y)$ for all $x \in \tilde{S}$.

The dynamics, i.e., terms like $\mathbb{P}^x(\tau_y^+ < \tau_x^+)$ or $\mathbb{P}^x(X_{\tau^+} = y)$ are usually hard to compute, and so will be the two transformations that are defined via the dynamics, but Lemma 8 below shows that approximating them is safe as far as the stable states are concerned.

Lemma 8 *Let p and \tilde{p} be perturbations with the same state space, such that $x \neq y \Rightarrow p(x, y) \cong \tilde{p}(x, y)$. For all stationary distribution maps μ for p , there exists $\tilde{\mu}$ for \tilde{p} such that $\mu \cong \tilde{\mu}$.*

E.g., both coefficients in Figure 2b (6b) can safely be replaced with $\epsilon(1)$, and Figure 4b can be replaced with Figure 4c. Lemma 8 will dramatically simplify the computation of the stable states.

2.1 Essential graph

The *essential graph* of a perturbation captures the non-infinitesimal flow between different states at the normal time scale. It is a very coarse description of the perturbation.

Definition 9 (Essential graph) *Given a perturbation with state space S , the essential graph is a binary relation over S and possesses the arc (x, y) if $x \neq y$ and $p(z, t) \precsim p(x, y)$ for all $z, t \in S$. The essential classes are the sink (aka bottom) strongly connected components of the graph. The other SCCs are the transient classes. A state in an essential class is essential, the others are transient.*

The essential classes will be named E_1, \dots, E_k . Observation 10 below implies that the essential graph is made of the arcs (x, y) such that $x \neq y$ and $p(x, y) \cong 1$, as expected.

Observation 10 *Let p be a perturbation. There exist positive c and ϵ_0 such that for all $\epsilon < \epsilon_0$, for all simple paths γ in the essential graph, $c < p_\epsilon(\gamma)$.*

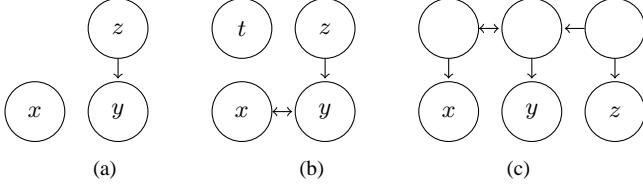


Figure 3: Essential graphs

For example, the perturbations (with $I =]0, 1]$) that are described in Figures 1b, 1c, 2a, and 2c all have Figure 3a as essential graph, and $\{x\}$ and $\{y\}$ as essential class. Figure 3b (3c) is the essential graph of Figure 4a (5a), and $\{x, y\}$ and $\{t\}$ are its essential classes. Note that the essential states of a perturbation and the essential states of a Markov chain are two distinct (yet related) concepts: *e.g.*, all states from Figure 4a are essential for the Markov chain for all $\epsilon \in]0, 1]$.

The essential graph alone cannot tell which states are stable: *e.g.*, swapping ϵ and ϵ^2 in Figure 2a yields the same essential graph but Lemma 6 shows that the only stable state is then x instead of y . The graph allows us to make the following case disjunction nonetheless, along which we will either say that all states are stable, or perform one of the transformations from the next subsections.

1. Either the graph is empty (*i.e.* totally disconnected) and the perturbation is zero, or
2. it is empty and the perturbation is non-zero, or
3. it is non-empty and has a non-singleton essential class, or
4. it is non-empty and has only singleton essential classes.

Observation 10 motivates the following convenient assumption.

Assumption 11 *There exists $c > 0$ such that $p(\gamma) > c$ for every simple path γ in the essential graph.*

The two assumptions above do not have the same status: Assumption 5 is a key condition that will appear explicitly in our final result, whereas Assumption 11 is just made wlog, *i.e.*, up to focusing on a smaller neighborhood of 0 inside I .

Lemma 12 shows the usefulness of Assumption 11. It is proved by Lemma 6, and is used later to strengthen Lemma 7.2 into $\mu \cong \tilde{\mu}$.

Lemma 12 *Let a perturbation p with state space S and transient states T satisfy Assumption 11. Then $\frac{c}{c+|S|} \leq \sum_{x \in S \setminus T} \mu(x)$.*

2.2 Essential collapse

The essential collapse, defined below, amounts to merging one essential class of a perturbation into one meta-state and letting this state represent faithfully the whole class in terms of dynamics between the whole class and each of the outside states.

Definition 13 (Essential collapse of a perturbation) *Let p be a perturbation on state space S . Let x be a state in an essential class E , and let $\tilde{S} := (S \setminus E) \cup \{\cup E\}$. The essential collapse $\kappa(p, x) : I \times \tilde{S} \times \tilde{S} \rightarrow [0, 1]$ of p around x is defined below.*

$$\begin{aligned} \kappa(p, x)(\cup E, \cup E) &:= \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}}^+ = x) \\ \kappa(p, x)(\cup E, y) &:= \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}}^+ = y) \quad \text{for all } y \in S \setminus E \\ \kappa(p, x)(y, \cup E) &:= \sum_{z \in E} p(y, z) \quad \text{for all } y \in S \setminus E \\ \kappa(p, x)(y, z) &:= p(y, z) \quad \text{for all } y, z \in S \setminus E \end{aligned}$$

Observation 14 $\kappa(p, x)$ is again a perturbation, κ preserves irreducibility, and if $\{x\}$ is an essential class, $\kappa(p, x) = p$.

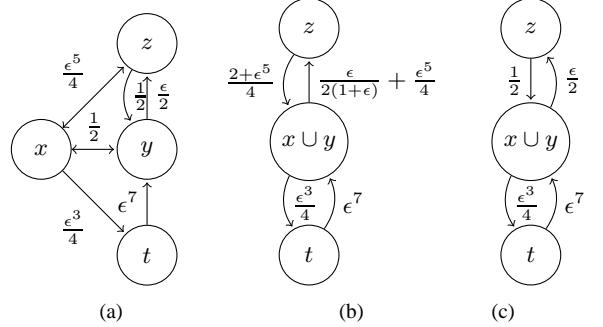


Figure 4: Essential collapse

For example, collapsing around x or y in Figure 2b has no effect. The perturbation in Figure 4a has two essential classes, *i.e.*, its essential graph has two sink SCCs, namely $\{x, y\}$ and $\{t\}$. Figure 4b displays its essential collapse around x . It was calculated by noticing that $\mathbb{P}^x(X_{\tau_{\{x,z,t\}}^+} = t) = \frac{\epsilon^3}{4}$, and $\mathbb{P}^x(X_{\tau_{\{x,z,t\}}^+} = x) = \frac{1}{2} - \frac{\epsilon^3}{4} - \frac{\epsilon^5}{4} + \frac{1}{2} \cdot \mathbb{P}^y(X_{\tau_{\{x,z,t\}}^+} = x)$, and $\mathbb{P}^y(X_{\tau_{\{x,z,t\}}^+} = x) = \frac{1}{2} + \frac{1-\epsilon}{2} \cdot \mathbb{P}^y(X_{\tau_{\{x,z,t\}}^+} = x)$.

Proposition 18 will show that it suffices to compute the stable states of Figure 4b to compute those of Figure 4a, and by Lemma 8 it suffices to compute those of the simpler Figure 4c. However, computing the exact values $\mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}^+} = y)$ can be difficult even on simple examples like above. Fortunately, Lemma 15 shows that they are \cong -equivalent to maxima that are easy to compute. *E.g.*, using Lemma 15 to approximate the essential collapse of Figure 4a around x yields Figure 4c, but without having to compute the intermediate Figure 4b.

Lemma 15 *Let p be a perturbation with state space S satisfy Assumption 5, and let \tilde{p} be the essential collapse $\kappa(p, x)$ of p around x in some essential class E . For all $y \in S \setminus E$, we have $\tilde{p}(\cup E, y) \cong \max_{z \in E} p(z, y)$ and $\tilde{p}(y, \cup E) \cong \max_{z \in E} p(y, z)$.*

Note that by Lemma 15, only the essential class is relevant during the essential collapse up to \cong , the exact state is irrelevant. Lemma 15 is also a tool that is used to prove, *e.g.*, Proposition 16 below which shows that the essential graph may contain useful information about the stable states.

Proposition 16 *Let a perturbation p with state space S satisfy Assumption 5, let μ be a corresponding stationary distribution map.*

1. If y is a transient state, $\liminf_{\epsilon \rightarrow 0} \mu_\epsilon(y) = 0$.
2. If two states x and y belong to the same essential or transient class, $\mu(x) \cong \mu(y)$.

Proposition 16.1 says that the transient states are vanishing, *e.g.* the nameless states in Figure 3c. Proposition 16.2 says that two states in the same class are either both stable or both vanishing, *e.g.* $\{x\}$ and $\{y\}$ in Figure 3b.

The usefulness of the essential collapse comes from its preserving and reflecting stability, as stated in Proposition 18. Its proof invokes Lemma 17 below, which shows that the essential collapse preserves the dynamics up to \cong , and Lemma 6, which relates the dynamics and the stationary distributions.

Lemma 17 *Given a perturbation p with state space S , let \tilde{p} be the essential collapse of p around x in some essential class E , and let $\tilde{x} := \cup E$. The following holds for all $y \in S \setminus E$.*

$$\mathbb{P}^y(\tau_x < \tau_y) \cong \tilde{\mathbb{P}}^y(\tau_{\tilde{x}} < \tau_y) \quad \wedge \quad \mathbb{P}^x(\tau_y < \tau_x) \cong \tilde{\mathbb{P}}^x(\tau_y < \tau_x)$$

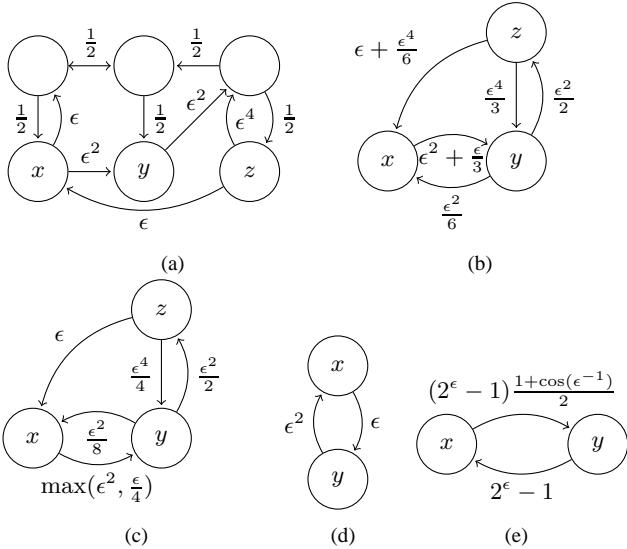


Figure 5: Transient deletion (mainly)

Proposition 18 Let a perturbation p with state space S satisfy Assumption 5, and let x be in an essential class E .

1. Let \tilde{p} be the chain after the essential collapse of p around x . Let $\mu(\tilde{p})$ be a stationary distribution map of $p(\tilde{p})$. There exists a stationary distribution map $\tilde{\mu}$ for \tilde{p} (μ for p) such that $\tilde{\mu}(\cup E) \cong \mu(x)$ and $\tilde{\mu}(y) \cong \mu(y)$ for all $y \in S \setminus E$.
2. A state $y \in S$ is stable for p iff either $y \in E$ and $\cup E$ is stable for $\kappa(p, x)$, or $y \notin E$ and y is stable for $\kappa(p, x)$.

By definition, collapsing an essential class preserves the structure of the perturbation outside of the class, so Proposition 18 implies that the essential collapse commutes up to \cong . Especially, the order in which the essential collapses are performed is irrelevant as far as the stable states are concerned.

2.3 Transient deletion

If all the essential classes of a perturbation are singletons, Observation 14 says that the essential collapse is useless. If in addition the essential graph has arcs, there are transient states, and Definition 19 below deletes them to shrink the perturbation further.

Definition 19 (Transient deletion) Let a perturbation p with state space S , transient states T , and singleton essential classes, satisfy Assumption 5. The function $\delta(p)$ over $S \setminus T$ is derived from p by transient deletion: for all distinct $x, y \in S \setminus T$ let

$$\delta(p)(x, y) := \mathbb{P}^x(X_{\tau_{S \setminus T}^+} = y)$$

Observation 20 $\delta(p)$ is again a perturbation, δ preserves irreducibility, and if all states are essential, $\delta(p) = p$.

For example, in Figure 2a the essential classes are $\{x\}$ and $\{y\}$, z is transient, and the transient deletion yields Figure 5d. Also, in Figure 5a, the essential classes are $\{x\}$, $\{y\}$, and $\{z\}$, the transient states are nameless, and the transient deletion yields Figure 5b.

The transient deletion is useful thanks to Proposition 21 below, whose proof relies on Lemmas 7.2 and 12.

Proposition 21 If a perturbation p satisfy Assumption 5 and has singleton essential classes, p and $\delta(p)$ have the same stable states.

Like the essential collapse, the transient deletion is defined via the dynamics and is hard to compute. Like Lemma 15 did for the

essential collapse, Lemma 22 approximates the transient deletion by an expression that is easy to compute.

Lemma 22 If a perturbation p with state space S and transient states T satisfies Assumption 5 and has singleton essential classes, $\mathbb{P}^x(X_{\tau_{S \setminus T}^+} = y) \cong \max\{p(\gamma) : \gamma \in \Gamma_T(x, y)\}$ for all $x, y \in S \setminus T$.

E.g., Figure 5a yields Figure 5c without computing Figure 5b. Note that $\max(\epsilon^2, \frac{\epsilon}{4})$ in Figure 5c may be simplified into ϵ by Lemma 8.

2.4 Outgoing scaling and existence of stable states

If the essential graph has no arc, the essential collapse and the transient deletion are useless to compute the stable states. This section says how to transform a non-zero perturbation with empty (i.e. totally disconnected) essential graph into a perturbation with the same stable states but a non-empty essential graph, so that collapse or deletion may be applied. Roughly speaking, it is done by speeding up time until a first non-infinitesimal flow is observable between different states, i.e. until the new essential graph has arcs.

Towards it, the *ordered division* is defined in Definition 23. It allows us to divide a function by a function with zeros by returning a default value in the zero case. It is named ordered because we will "divide" f by g only if $f \precsim g$, so that only 0 may be "divided" by 0. Then Observation 24 further justifies the terminology.

Definition 23 (Ordered division) For $f, g : I \rightarrow [0, 1]$ and $n > 1$ let us define $(f \div_n g) : I \rightarrow [0, 1]$ by $(f \div_n g)(x) := \frac{f(x)}{g(x)}$ if $0 < g(x)$ and otherwise $(f \div_n g)(x) := \frac{1}{n}$.

Observation 24 $(f \div_n g) \cdot g = f$ for all n and $f, g : I \rightarrow [0, 1]$ such that $f \precsim g$.

Definition 25 (Outgoing scaling) Let a perturbation p with state space S satisfy Assumption 5, let $m := |S| \cdot \max\{p(z, t) \mid z, t \in S \wedge z \neq t\}$, and let us define the following.

- $\sigma(p)(x, y) := p(x, y) \div_{|S|} m$ for all $x \neq y$
- $\sigma(p)(x, x) := (p(x, x) + m - 1) \div_{|S|} m$.

For example, Figure 2b satisfies Assumption 5 and its essential graph is empty, i.e. totally disconnected. Applying outgoing scaling to it yields Figure 6b, which satisfies Assumption 5 and whose essential graph has two arcs. Note that collapsing around x or y in Figure 2b has no effect, but in Figure 6b it yields a one-state perturbation. Also, Figure 6a does not satisfy Assumption 5 and its essential graph is empty. Applying outgoing scaling to it yields Figure 6c, which does not satisfy Assumption 5 and whose essential graph has one arc. Applying it again to Figure 6c would only divide the non-self-loop coefficients by 3. More generally, Proposition 26 below states how well the outgoing scaling behaves.

Proposition 26 1. If a perturbation p satisfies Assumption 5, so does $\sigma(p)$, and the essential graph of $\sigma(p)$ is non-empty .
2. A state is stable for p iff it is stable for $\sigma(p)$.

The outgoing scaling divides the weights of the proper arcs by m , as if time were sped up by m^{-1} . The self-loops thus lose their meaning, but Proposition 26 proves it harmless. Note that the self-loops are also ignored in Assumption 5, Lemma 8, and Definition 9.

Let us now describe a recursive procedure computing the stable states: if the perturbation is zero, all its states are stable; else, if the essential graph is empty, apply the outgoing scaling; else, apply one essential collapse or the transient deletion. This procedure is correct by Propositions 26.2, 18.2, and 21, hence Theorem 27 below, which is the existential part of Theorem 2.

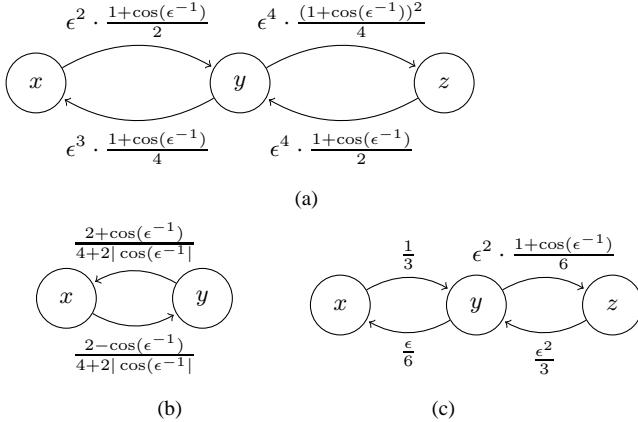


Figure 6: Outgoing scaling

Theorem 27 Let p be a perturbation such that $f \precsim g$ or $g \precsim f$ for all f and g in the multiplicative closure of the $p(x, y)$ with $x \neq y$. Then p has stable states.

3. Abstract and quick algorithm

The procedure described before Theorem 27 computes the stable states, but a very rough analysis of its algorithmic complexity shows that it runs in $O(n^7)$, where n is the number of states. (A better analysis might find $O(n^5)$.) This bad complexity comes from the difficulty to analyze the procedure precisely and from some redundant operations done by the transformations, especially the successive essential collapses. Instead we will perform the successive collapses followed by one transient deletion as a single transformation. Applying alternately the outgoing scaling and the new transformation, both up to \cong , is the base of our algorithm.

Section 3.1 abstracts the relevant notions up to \cong and gives useful algebraic properties that they satisfy. Based on these abstractions, Section 3.2 presents the algorithm (computing the stable states and more), its correctness, and its complexity in $O(n^3)$.

3.1 Abstractions

Ensuring that the essential collapse and the transient deletion can be safely performed up to \cong is a straightforward sanity check, by Lemma 8. However, the proof for the outgoing scaling involves a new algebraic structure to accommodate the ordered division, and handling the combination of the successive collapses and one deletion requires particular attention. It would have been cumbersome to define this combination directly *via* the dynamics in Section 2, and more difficult to prove its correctness *via* probabilistic techniques, hence the usefulness of the rather atomic collapse and deletion.

Our first step below is to consider functions up to \cong .

Definition 28 (Equivalence classes and quotient set) For $f : I \rightarrow [0, 1]$ let $[f]$ be its \cong equivalence class; for a matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with elements in $I \rightarrow [0, 1]$, let $[A]$ be the matrix where $[A]_{ij} := [a_{ij}]$ for all $0 \leq i, j \leq n$. For a set F of functions from I to $[0, 1]$, let $[F]$ be the quotient set F / \cong . Finally, it is possible to lift over $[F]$ both \cdot to $[\cdot]$ and \precsim to $[\precsim]$.

Observation 29 For (G, \cdot) a semigroup totally preordered by \precsim ,

1. $[\precsim]$ orders $[G]$ linearly, so $\max_{[\precsim]}$ is well-defined.
2. $([G] \cup \{[\epsilon \mapsto 0]\}, [\epsilon \mapsto 0], [\epsilon \mapsto 1], \max_{[\precsim]}, [\cdot])$ is a commutative semiring. (See, e.g., [8] for the related definitions.)

The good behavior of \cdot and \precsim up to \cong is expressed above within an existing algebraic framework, but for \div_n we introduce a new algebraic structure below.

Definition 30 (Ordered-division semiring) An ordered-division semiring is a tuple $(F, 0, 1, \cdot, \leq, \div)$ such that (F, \leq) is a linear order with maximum 1, and $(F, 0, 1, \max_{\leq}, \cdot)$ is a commutative semiring, and for all $f \leq g$ we have $f \div g$ is in F and $(f \div g) \cdot g = f$.

Observation 31 Let $(F, 0, 1, \cdot, \div, \leq)$ be an ordered-division semiring. Then $0 = \min_{\leq} F$ and $f \div 1 = f$ for all f .

Lemma 32 below shows that the functions $I \rightarrow [0, 1]$ up to \cong form an ordered-division semiring.

Lemma 32 1. Let $n > 1$ and $f, f', g, g' : I \rightarrow [0, 1]$ be such that $f \cong f' \precsim g \cong g'$. Then $[f \div_1 g] = [f' \div_n g']$, which we then write $[f][\div][g]$.

2. For all sets G of functions from I to $[0, 1]$ closed under multiplication, the tuple $([G \cup \{\epsilon \mapsto 0\}], [\epsilon \mapsto 0], [\epsilon \mapsto 1], [\cdot], [\div], [\precsim])$ is an ordered-division semiring.

For example, the set containing $[\epsilon \mapsto 0]$ and all the $[\epsilon \mapsto \epsilon^\alpha]$ for non-positive α is an ordered-division semiring, where $[\epsilon \mapsto \epsilon^\alpha][\cdot][\epsilon \mapsto \epsilon^\beta] = [\epsilon \mapsto \epsilon^{\alpha+\beta}]$ and $[\epsilon \mapsto \epsilon^\alpha][\precsim][\epsilon \mapsto \epsilon^\beta]$ iff $\beta \leq \alpha$, and $[\epsilon \mapsto 0][\precsim][\epsilon \mapsto \epsilon^\alpha]$, and $[\epsilon \mapsto \epsilon^\alpha][\div][\epsilon \mapsto \epsilon^\beta] = [\epsilon \mapsto \epsilon^{\alpha-\beta}]$ for $\beta \leq \alpha$.

To handle σ , κ , and δ up to \cong we define below transformations of weighted graphs with weights in an ordered-division semiring.

Definition 33 (Abstract transformations) Let $P : S \times S \rightarrow F$, where $(F, 0, 1, \cdot, \leq, \div)$ is an ordered-division semiring.

1. Let $\{(z, t) \in S^2 \mid P(z, t) = 1 \wedge z \neq t\}$ be the essential graph of P , and let the sink SCCs E_1, \dots, E_k be its essential classes.
2. **Outgoing scaling:** for $x \neq y$ let $[\sigma](P)(x, y) := P(x, y) \div M$, where $M := \max_{\leq} \{P(z, t) : (z, t) \in S \times S \wedge z \neq t\}$, and $[\sigma](P)(x, x) := 1$.
3. **Essential collapse:** let $[\kappa](P, E_i)$ be the matrix with states $\{\cup E_i\} \cup S \setminus E_i$ such that for all $x, y \in S \setminus E_i$ we set $[\kappa](P, E_i)(x, y) := P(x, y)$ and $[\kappa](P, E_i)(\cup E_i, y) := \max_{\leq} \{P(x_i, y) : x_i \in E_i\}$ and $[\kappa](P, E_i)(x, \cup E_i) := \max_{\leq} \{P(x, x_i) : x_i \in E_i\}$ and $[\kappa](P, E_i)(\cup E_i, \cup E_i) := 1$.
4. **Shrinking:** let $[\chi](P)$ be the matrix with state space $\{\cup E_1, \dots, \cup E_k\}$ such that for all i, j $[\chi](P)(\cup E_i, \cup E_j) := \max_{\leq} \{P(\gamma) : \gamma \in \Gamma_T(E_i, E_j)\}$.

In Definition 33, the weights $P(x, x)$ occur only in the definitions of the self-loops of the transformed graphs, whence Observation 34 below.

Observation 34 Let $(F, 0, 1, \cdot, \leq, \div)$ be an ordered-division semiring, let $P, P' : S \times S \rightarrow F$ be such that $P(x, y) = P'(x, y)$ for all $x \neq y$. Then P and P' have the same essential graph and classes E_1, \dots, E_k ; $[\sigma](P)(x, y) = [\sigma](P')(x, y)$ for all $x \neq y$; and for all l and $i \neq j$ we have $[\chi](P)(\cup E_i, \cup E_j) = [\chi](P')(\cup E_i, \cup E_j)$ and $[\kappa](P)(P, E_l)(\cup E_i, \cup E_j) = [\kappa](P')(P', E_l)(\cup E_i, \cup E_j)$.

Lemma 35 below shows that the transformations from Definition 33 are faithful abstractions of σ , κ , and δ . Some proofs come with examples, which also highlight the benefits of abstraction.

Lemma 35 (Abstract and concrete transformations) Let a perturbation p with state space S satisfy Assumption 5, let E_1, \dots, E_k be its essential classes, and for all i let $x_i \in E_i$.

1. p and $[p]$ have the same essential graph.
2. $[\sigma]([p])(x, y) = [\sigma(p)](x, y)$ for all $x \neq y$.

3. $[\chi]([p])(\{x\}, \{y\}) = [\delta(p)](x, y)$ whenever $\delta(p)$ is well-defined.
4. $[\kappa]([p], E_i) = [\kappa(p, x_i)]$.
5. $[\chi]([p]) = [\chi] \circ [\kappa]([p], E_1)$.
6. $[\delta \circ \kappa(\dots \kappa(\kappa(\sigma(p), x_1), x_2), \dots, x_k)](\cup E_i, \cup E_j) = [\chi] \circ [\sigma]([p])(\cup E_i, \cup E_j)$ for all $i \neq j$.

By the algorithm underlying Theorem 27 and Lemma 35.6 we are now able to state the following.

Proposition 36 *Let a perturbation p satisfy Assumption 5. There exists $n \in \mathbb{N}$ such that $([\chi] \circ [\sigma])^n([p])(x, y) = 0$ for all $x \neq y$ in its state space. Furthermore, the states of $([\chi] \circ [\sigma])^n([p])$ correspond to the stable states of p .*

3.2 The algorithm

Algorithm 1 mainly consists in applying recursively the function $[\chi] \circ [\sigma]$ occurring in Proposition 36 until an empty (*i.e.* totally disconnected) graph is produced. It does not explicitly refer to perturbations since this notion was abstracted on purpose. Instead, the algorithm manipulates digraphs with arcs labeled in an ordered-division semiring, in which inequality, multiplication and ordered division are implicitly assumed to be computable.

One call to the recursive function HubRec corresponds to $[\chi] \circ [\sigma]$, *i.e.* Lines 7 and 9 correspond to $[\sigma]$, and Lines 10 till 18 correspond to $[\chi]$. Before calling HubRec Lines 2 and 3 produce an isomorphic copy of the input, which will be easier to handle when making unions and keeping track of the remaining vertices. Note that Line 9 does not update the $P(x, x)$. It would be useless indeed, since in Definition 33 the self-loops of the original graph occur only in the definition of the self-loops of the transformed graphs, and since self-loops are irrelevant by Observation 34. Line 10 computes the essential graph, up to self-loops, and Line 11 computes the essential classes by a modified Tarjan's algorithm, as detailed in Algorithm 2. The computation of $[\chi](P)(\cup E_i, \cup E_j) := \max_{\leq} \{P(\gamma) : \gamma \in \Gamma_T(E_i, E_j)\}$ is performed in two stages: the first stage at Line 12 considers only paths of length one; the second stage at Line 18 considers only paths of length greater than one, and therefore having their second vertex in T . This case disjunction reduces the size of the graph on which the shortest path algorithm from Line 16 is run, and thus reduces the complexity of Hub from $O(n^4)$ to $O(n^3)$, as will be detailed in Proposition 37. Note that the shortest path algorithm is called with laws \cdot and \max instead of $+$ and \min . Moreover, since our weights are at most 1 we can use [14] or [2] (which assume non-negative weights) to implement Line 16.

Proposition 37 below shows that our algorithm is fast.

Proposition 37 *The algorithm Hub terminates within $O(n^3)$ computation steps, where n is the number of vertices of the input graph.*

By Propositions 36 and 37 we now state our main algorithmic result, which is the algorithmic part of Theorem 2.

Theorem 38 *Let a perturbation p satisfy Assumption 5. A state is stochastically stable iff it belongs to $\text{Hub}(S, [p])$. Provided that inequality, multiplication, and ordered division between equivalence classes of perturbation maps can be computed in constant time, stability can be decided in $O(n^3)$, where n is the number of states.*

One of the achievements of our algorithm is that it processes all weighted digraphs (*i.e.* abstractions of perturbations) uniformly. Especially, neither irreducibility nor any kind of connectedness is required. For example in Figure 7, the four-state perturbation is the disjoint union of two smaller perturbations. As expected the stable states of the union are the union of the stable states, *i.e.* $\{x, z\}$, but whereas the outgoing scaling applied to the bottom of

```

1 Function Hub is
input :  $(S, P)$ , where  $P : S \times S \rightarrow F$ 
//  $(F, 0, 1, \cdot, \leq, \div)$  is an ordered-division
semiring.
output: a subset of  $S$ 
2  $\hat{S} \leftarrow \{\{s\} | s \in S\};$  // For bookkeeping.
3 for  $x, y \in S$  do  $\hat{P}(\{x\}, \{y\}) \leftarrow P(x, y);$  // For
bookkeeping.
4 return HubRec( $\hat{S}, \hat{P}$ );
5 end

6 Function HubRec is
input :  $(S, P)$ , where  $S$  is a set of sets and
 $P : S \times S \rightarrow F$ 
output: a subset of  $S$ 
7  $M \leftarrow \max\{P(x, y) | (x, y) \in S \times S \wedge x \neq y\};$ 
8 if  $M = 0$  then return  $\cup S;$  // Recursion base case
9 for  $x, y \in S$  and  $x \neq y$  do  $P(x, y) \leftarrow P(x, y) \div M;$ 
// Outgoing scaling.
10  $A \leftarrow \{(x, y) \in S \times S | P(x, y) = 1\};$  //  $A$  is a
digraph.
11  $(E_1, \dots, E_k) \leftarrow \text{TarjanSinkSCC}(S, A);$  // Returns
the sink SCCs of  $A$ .
// Maximal labels of direct arcs, below.
12 for  $1 \leq i, j \leq k$  do
 $P'(\cup E_i, \cup E_j) \leftarrow \max\{P(x, y) | (x, y) \in E_i \times E_j\};$ 
// Maximal labels of all relevant paths, in
the remainder.
13  $T \leftarrow S \setminus (E_1 \cup \dots \cup E_k);$ 
14  $P_T \leftarrow P;$  // Initialisation.
15 for  $(x, y) \in (S \setminus T) \times S$  do  $P_T(x, y) \leftarrow 0;$  // Drops
arcs not starting in  $T$ .
16 for  $y \in T$  do  $P_T(y, \_) \leftarrow \text{Dijkstra}(S, P_T, y, \cdot, \max);$ 
//  $P_T(y, \_)$  is the "distance" function from
 $y \in T$ , using  $\cdot$  and  $\max$ .
17 for  $1 \leq i, j \leq k$  and  $i \neq j$  and
 $(x_i, x_j, y) \in E_i \times E_j \times T$  do
18      $P'(\cup E_i, \cup E_j) \leftarrow$ 
 $\max(P'(\cup E_i, \cup E_j), P(x_i, y) \cdot P_T(y, x_j));$ 
19 end for
20 return HubRec( $\{\cup E_1, \dots, \cup E_k\}, P'$ )
21 end

```

Algorithm 1: Hub

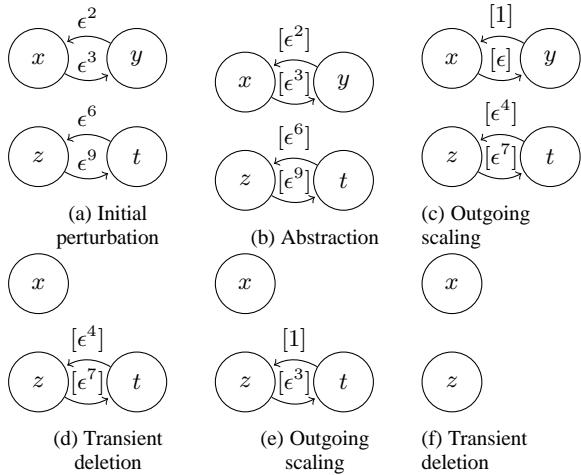


Figure 7: The algorithm run on a disconnected perturbation

Figure 7b (the perturbation restricted to $\{z, t\}$) would yield the bottom of Figure 7e directly by division by $[\epsilon^6]$, two rounds of outgoing scaling lead to this stage when processing the four-state perturbations.

4. Discussion

This section studies two special cases of our setting: first, how assumptions that are stronger than Assumption 5 make not only some proofs easier but also one result stronger; second, how far Young's technique can be generalized. Then we notice that the termination of our algorithm defines an induction proof principle, which is used to show that the algorithm computes a well-known object when fed a strongly connected graph. Eventually, we discuss how to give the so-far-informal notion of time scale a formal flavor.

4.1 Stronger assumption

Let us consider Assumption 39, which is stronger version of Assumption 5. Assumption 39 yields Proposition 40, which is stronger version of Proposition 16.1. (The proofs are similar but the new one is simpler.)

Assumption 39 *If $x \neq y$ and $p(x, y)$ is non-zero, it is positive; and $f \cong g$ or $f \in o(g)$ or $g \in o(f)$ for all f and g in the multiplicative closure of the $\epsilon \mapsto p_\epsilon(x, y)$ with $x \neq y$.*

Proposition 40 *Let a perturbation p with state space S satisfy Assumption 39, and let μ be a stationary distribution map for p . If y is a transient state, $\lim_{\epsilon \rightarrow 0} \mu_\epsilon(y) = 0$.*

Under Assumption 5 some states may be neither stable nor fully vanishing: y in Figure 5e and x in Figure 1a where the bottom ϵ^2 is replaced with ϵ . Assumption 39 rules out such cases, as below.

Corollary 41 *If a perturbation p satisfies Assumption 39, every state is either stable or fully vanishing.*

4.2 Generalization of Young's technique

Our approach to prove the existence of and compute the stable states of a perturbation is different from Young's approach [19] which uses a finite version of the Markov chain tree theorem. In this section we investigate how far Young's technique can be generalized. This will suggest that we were right to change approaches, but it will also yield a decidability result in Proposition 45.

Lemma 42 below is a generalization of [19, Lemma 1]. Both proofs use the Markov chain tree theorem, but they are significantly different nonetheless. Let p be a perturbation with state space S . As in [19] or [8], for all $x \in S$ let \mathcal{T}_x be the set of the spanning trees of (the complete graph of) $S \times S$ that are directed towards x . For all $x \in S$ let $\beta_x^\epsilon := \max_{T \in \mathcal{T}_x} \prod_{(z,t) \in T} p_\epsilon(z, t)$.

Lemma 42 *A state x of an irreducible perturbation with state space S is stable iff $\beta^y \precsim \beta^x$ for all $y \in S$.*

Assumption 5 and Lemma 42 together yield Observation 43, a generalization of existing results about existence of stable states, such as [19, Theorem 4]. The underlying algorithm runs in time $O(n^3)$ where n is the number of states, just like Young's.

Observation 43 *Let a perturbation p on state space S satisfy Assumption 5. If for all $x \neq y$ the map $p(x, y)$ is either identically zero or strictly positive, p has stable states.*

The stable states of a perturbation are computable even without the positivity assumption from Observation 43, but their existence is no longer guaranteed by the same proof. In this way, Observation 44 is like the existential part of Theorem 2, but with a bad complexity.

Observation 44 *Let F be a set of perturbation maps of type $I \rightarrow [0, 1]$ for some I . Let us assume that F is closed under multiplication by elements in F and by characteristic functions of decidable subsets of I , that \precsim is decidable on $F \times F$, and that the supports of the functions in F are uniformly decidable. If $f \precsim g$ or $g \precsim f$ for all $f, g \in F$, stability is decidable in $O(n^5)$ for the perturbations p such that $x \neq y \Rightarrow p(x, y) \in F$.*

The assumption $f \precsim g$ or $g \precsim f$ for all $f, g \in F$ from Observation 44 corresponds to Assumption 5. Proposition 45 below drops it while preserving decidability of stability, but at the cost of an exponential blow-up because the supports of the maps are no longer ordered by inclusion.

Proposition 45 *Let F be a set of perturbation maps of type $I \rightarrow [0, 1]$ for some I . Let us assume that F is closed under multiplication by elements in F and by characteristic functions of decidable subsets of I , that \precsim is decidable on $F \times F$, and that the supports of the functions in F are uniformly decidable. Then stability is decidable for the perturbations p such that $x \neq y \Rightarrow p(x, y) \in F$.*

4.3 What does Algorithm 1 compute?

Applying sequentially outgoing scaling, essential collapse, and transient deletion terminates. So it amounts to an *induction proof principle* for finite graphs with arcs labeled in an ordered-division semiring. Observation 46 is proved along this principle. It can also be proved by a very indirect argument using Lemma 42 and Theorem 38, but the proof using induction is simple and from scratch.

Observation 46 *Let $(F, 0, 1, \cdot, \leq, \div)$ be an ordered-division semiring, and let $P : S \times S \rightarrow F$ correspond to a strongly connected digraph, where an arc is absent iff its weight is 0. Then $\text{Hub}(S, P)$ returns the roots of the maximum directed spanning trees.*

Note that finding the roots from Observation 46 is also doable in $O(n^3)$ by computing the maximum spanning trees rooted at each vertex, by [7] which uses the notion of *heap*, whereas Hub uses a less advanced algorithm.

Observation 46 may be extended to non strongly connected digraphs by considering the sink SCCs independently, but alternatively it is not obvious how to generalize the notion of maximal spanning tree into a notion that is meaningful for non-strongly connected graphs. Nevertheless, the vertices returned by $\text{Hub}(S, P)$ are

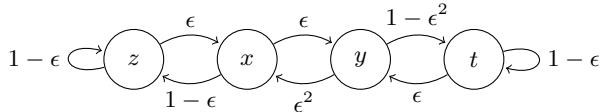


Figure 8: Vanishing time scale

the one in S that somehow attract the more flow/traffic according to P , hence the name Hub.

One last algorithmic remark: from the proof of Proposition 37 we see that Tarjan’s algorithm is an overkill to get a complexity of $O(n^3)$. Indeed, combining several basic shortest path-algorithms would have done the trick, but using Tarjan’s algorithm should make the computation of Hub faster by a constant factor.

4.4 Vanishing time scales

Under Assumption 5, computing Hub and considering the intermediate weighted graphs shows the order in which the states are found to be vanishing. Under the stronger Assumption 39, a notion of *vanishing time scale* may be defined, with the flavor of non-standard analysis [15]. Let (\mathcal{T}, \cdot) be a group of functions $I \rightarrow]0, +\infty[$ such that $f \cong g$ or $f \in o(g)$ or $g \in o(f)$ for all f and g in \mathcal{T} . The elements of $[\mathcal{T}]$ are called the time scales. Let a perturbation p on state space S satisfy Assumption 39 and let $x \in S$ be deleted at the d -th recursive call of $\text{Hub}(S, [p])$. Let M_1, \dots, M_d be the maxima (i.e. M) from Line 7 in Algorithm 1 at the 1st,..., d -th recursive calls, respectively. We say that x vanishes at time scale $\prod_{1 \leq i \leq d} M_i^{-1}$.

Figure 8 suggests that a similar account of vanishing times scales, even just a qualitative one, would be much more difficult to obtain by invoking the Markov chain tree theorem as in [19]. The only stable states is t ; the state z vanishes at time scale $[\epsilon]^{-2}$; and x and y vanish at the same time scale [1] although the maximum spanning trees rooted at x and y have different weights: ϵ^4 and ϵ^3 , respectively.

Acknowledgments

We thank Ocan Sankur for useful discussions.

References

- [1] V. Betz and S. L. Roux. Multi-scale metastable dynamics and the asymptotic stationary distribution of perturbed Markov chains, 2014. <http://arxiv-web3.library.cornell.edu/abs/1412.6979v1>.
- [2] E. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959. ISSN 0029-599X. doi: 10.1007/BF01386390. URL <http://dx.doi.org/10.1007/BF01386390>.
- [3] G. Ellison. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies*, 67:17–45, 2000.
- [4] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107115, 1935.
- [5] D. Foster and P. Young. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232, 1990.
- [6] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*. Springer-Verlag, second edition, 1998.
- [7] H. Gabow, Z. Galil, T. Spencer, and R. Tarjan. Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. *Combinatorica*, 6:109–122, 1986.
- [8] B. B. Gursoy, S. Kirkland, O. Mason, and S. Sergeev. The Markov chain tree theorem in commutative semirings and the state reduction algorithm in commutative semifields. *Linear Algebra and its Applications*, 468:184–196, 2015. 18th ILAS Conference.
- [9] A. G. J.R. Wicks. An algorithm for computing stochastically stable distributions with applications to multiagent learning in repeated games. In *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, 2005. <http://arxiv.org/abs/1207.1424>.
- [10] M. Kandori, G. J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56, 1993.
- [11] H.-H. Kohler and E. Vollmerhaus. The frequency of cyclic processes in biological multistate systems. *J. Math. Biology*, 9:275–290, 1980.
- [12] H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284304, 1940.
- [13] F. Leighton and R. Rivest. The Markov chain tree theorem. Technical Report LCS-TM-249, MIT, 1983.
- [14] M. Leyzorek, R. Gray, A. Johnson, W. Ladew, S. Meaker, R. P. Jr, and R. Seitz. A study of model techniques for communication systems. Technical report, Case Institute of Technology, Cleveland, Ohio,, 1957. Investigation of Model Techniques, First Annual Report.
- [15] A. Robinson. *Non-standard Analysis*. Princeton University Press, 1974.
- [16] J. M. Smith and G. Price. The logic of animal conflicts. *Nature*, 246: 15–18, 1973.
- [17] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1:146–160, 1972.
- [18] Wikipedia. Tarjan’s strongly connected components algorithm — wikipedia, the free encyclopedia, 2015. [Online; accessed 25-March-2015].
- [19] P. Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.

A. Tarjan Modified

The function *TarjanSinkSCC* is written in Algorithm 2. It consists of Tarjan’s algorithm [17],[18], which normally returns all the SCCs of a directed graph, plus a few newly added lines (as mentioned in comments) so that it returns the sink SCCs only. It is not difficult to see that the newly added lines do not change the complexity of the algorithm, which is $O(|S| + |A|)$ where $|S|$ and $|A|$ are the numbers of vertices and arcs in the graph, respectively. The new lines only deal with the new boolean values $v.\text{sink}$. These lines are designed so that when popping an SCC with root v from the stack , the value $v.\text{sink}$ is true iff the SCC is a sink, hence the test at Line 35. All the $v.\text{sink}$ are initialized with *true* at Line 5, and $v.\text{sink}$ is set to false at two occasions: at Line 36 before a sink SCC with root v is output; and at Line 24 when one successor w of v has already been popped from the stack (since $w.\text{index}$ is defined), which means that there is one SCC below that of v . These are then propagated upwards at Line 19. The conjunction reflects the facts that a vertex is not in a sink SCC iff one of its successors in the same SCC is not.

B. Proofs and Lemmas

Lemma 47 below relates to Definition 3.

- Lemma 47** 1. \precsim is a preorder and \cong an equivalence relation.
2. For all $f, g : I \rightarrow]0, 1]$, we have $f \precsim g$ iff $\frac{1}{f} \precsim \frac{1}{g}$, so $f \cong g$ iff $\frac{1}{f} \cong \frac{1}{g}$.
3. $f \precsim g$ and $f' \precsim g'$ implies $f + f' \precsim g + g'$ and $f \cdot f' \precsim g \cdot g'$.
4. $f \cong g$ and $f' \cong g'$ and $f \precsim f'$ implies $g \precsim g'$.
5. $f + f' \cong \max(f, f') := x \mapsto \max(f(x), f'(x))$.
6. $f \precsim f'$ implies $\max(f, f') \cong f'$.
7. $f|_J \precsim g|_J$ and $f|_{I \setminus J} \precsim g|_{I \setminus J}$ implies $f \precsim g$.
8. Let 0 be a limit point of both $J \subseteq I$ and $I \setminus J$. A state x is stable (fully vanishing) for a perturbation p iff it is stable (fully vanishing) for both $p|_J$ and $p|_{I \setminus J}$.

```

1 Function TarjanSinkSCC is
2   input :  $(S, A)$ , where  $A \subseteq S \times S$ 
3   output: the sink SCC
4
5   index  $\leftarrow 0$ ;
6   stack  $\leftarrow \emptyset$ ;
7   for  $v \in S$  do  $v.onstack \leftarrow \text{false}$ ;
8   for  $v \in S$  do  $v.sink \leftarrow \text{true}$ ; // Newly added.
9   for  $v \in V$  do
10    | if  $v.index$  is undefined then StrongConnect( $v$ );
11    | end if
12
13   Function StrongConnect( $v$ ) is
14      $v.index \leftarrow index$ ;
15      $v.lowlink \leftarrow index$ ;
16      $index \leftarrow index + 1$ ;
17     stack.push( $v$ );
18      $v.onstack \leftarrow \text{true}$ ;
19
20     for  $(v, w) \in A$  do
21      | if  $w.index$  is undefined then
22        | | StrongConnect( $w$ );
23        | |  $v.lowlink \leftarrow$ 
24        | |  $\min(v.lowlink, w.lowlink)$ ;
25        | |  $v.sink \leftarrow v.sink \wedge w.sink$ ; // Newly
26        | | added.
27      | else
28        | | if  $w.onstack = \text{true}$  then
29          | | |  $v.lowlink \leftarrow$ 
30          | | |  $\min(v.lowlink, w.index)$ 
31        | | else
32          | | |  $v.sink \leftarrow \text{false}$ ; // Newly added.
33        | | end if
34      | end if
35    | end for
36
37    if  $v.lowlink = v.index$  then
38      | | start a new SCC;
39      | | repeat
40        | | |  $w \leftarrow stack.pop()$ ;
41        | | |  $w.onstack \leftarrow \text{false}$ ;
42        | | | add  $w$  to the current SCC;
43      | | until  $w = v$ ;
44
45    if  $v.sink$  then // Newly added.
46      | | |  $v.sink \leftarrow \text{false}$ ; // Newly added.
47      | | | output the SCC;
48    | end if
49  | end if
50 end
51 end

```

Algorithm 2: modification of Tarjan's SCC algorithm

Proof [Observation 4] Let $S := \{x_1, \dots, x_n, y_1, \dots, y_m\}$, for all $i < n - 1$ let $p(x_i, x_{i+1}) := f_i$, let $p(x_n, y_0) := f_n$, for all i let $p(x_i, x_1) := 1 - f_i$, for all $j < m - 1$ let $p(y_j, y_{j+1}) := g_j$, let $p(y_m, x_0) := g_m$, for all j let $p(y_j, y_1) := 1 - g_j$. It is easy to check that $\beta^{x_i} = \prod_j g_j \cdot \prod_{1 \leq k < i} f_i \cdot \prod_{i \leq k} \max(f_k, 1 - f_k) \cong \prod_j g_j \cdot \prod_{1 \leq k < i} f_i$ for all i . Let us first assume that the functions in F are positive. So by Lemma 42, if one x_i is stable, so is x_1 . Likewise for the y_j . But β^{x_1} and β^{y_1} are non comparable by assumption, so by Lemma 42 there are no stable state.

Let us now prove the claim in the general case, which holds if $n = m = 1$ for the same reason as in Figure 1a, so let us assume that $n + m > 2$. Wlog let us also assume that any two products of n' and m' functions from F with $n' + m' < n + m$ are comparable. So the functions in F are pairwise comparable; (up to intersection of I with a neighborhood of 0) their supports constitute a linearly ordered set for the inclusion; and $\prod_i f_i \lesssim \prod_{1 < j} g_j$ since $g_1 \leq 1$ and $\neg(\prod_j g_j \lesssim \prod_i f_i)$. Up to renaming let us assume that g_1 has the smallest support J . Up to restriction of I to the support of $\prod_i f_i$, which does not change stability nor non-comparability of $\prod_i f_i$ and $\prod_j g_j$ by Lemmas 47.8 and 47.7, let us assume that the f_i are positive, and so is g_j for $j > 1$ since $\prod_i f_i \lesssim \prod_{1 < j} g_j$. If g_1 is also positive, we are back to the special case above, so let us assume that it is not. On the one hand, restricting p to $I \setminus J$ shows that only y_1 might be stable, by Lemma 47.8; on the other hand $\neg(\prod_j g_j|_J \lesssim \prod_i f_i|_J)$, so let us make a case disjunction: if $\prod_i f_i|_J \lesssim \prod_j g_j|_J$ then $\beta^{y_1} \in o(\beta^{x_1})$, so y_1 cannot be stable by Lemma 42; if they are not comparable, the special case above says that the $p|_J$ has no stable state, so neither has p by Lemma 47.8. \square

Proof [Lemma 6] Let us assume that μ is stationary, so it is well-known that its support involves only essential states. To prove that the equation holds let us make a case disjunction: if x and y are transient states, $\mu(x) = \mu(y) = 0$; if x is transient and y is essential, $\mu(x) = 0$ and $\mathbb{P}^y(\tau_x^+ < \tau_y^+) = 0$; if x and y belong to distinct essential classes, $\mathbb{P}^x(\tau_y^+ < \tau_x^+) = \mathbb{P}^y(\tau_x^+ < \tau_y^+) = 0$; if x and y belong to the same essential class E , let E_1, \dots, E_k be the essential classes. Then for all $i \in \{1, \dots, k\}$ let μ_{E_i} be the extension to S (by the zero-function outside of E_i) of the unique stationary distribution of $p|_{E_i \times E_i}$. So $\frac{\mu(x)}{\mu(y)} = \frac{\mu_{E(x)}}{\mu_{E(y)}}$ by [1, Proposition 2.1] and since μ is a convex combination of the $\mu_{E_1}, \dots, \mu_{E_k}$.

Conversely, let us assume that the support of μ involves only essential states and that the equation holds. Let $E'_1, \dots, E'_{k'}$ be the essential classes with positive μ -measure. Let $i < k'$, and for all $x \in E'_i$ let $\mu_i(x) := \frac{\mu|_{E'_i}(x)}{\sum_{y \in E'_i} \mu(y)}$ define a distribution for $p|_{E'_i \times E'_i}$. Since μ_i also satisfies the equation and that $p|_{E'_i \times E'_i}$ is irreducible, μ_i is the unique stationary distribution for $p|_{E'_i \times E'_i}$. Since μ is a convex combination of the $\mu_1, \dots, \mu_{k'}$, it is stationary for p . \square

Proof [Lemma 7]

1. Let $\sigma_N := \sup\{n \in \mathbb{N} : |\{k \leq n : X_k \in \tilde{S}\}| = N\}$ be the supremum of the first time that $(X_n)_{n \in \mathbb{N}}$ has visited N states in \tilde{S} . Clearly $\sigma_N \xrightarrow{N \rightarrow \infty} \infty$, so $\mathbb{P}^x(\tau_y^+ < \tau_x^+) = \lim_{N \rightarrow \infty} \mathbb{P}^x(\tau_y^+ < \min(\tau_x^+, \sigma_N))$. On the other hand, $\mathbb{P}^x(\tau_y^+ < \min(\tau_x^+, \sigma_N))$

$$\begin{aligned}
&= \mathbb{P}^x(X_{\tau_{S \setminus T}^+} = y) \\
&+ \sum_{z \in S \setminus (T \cup \{x, y\})} \mathbb{P}^x(X_{\tau_{S \setminus T}^+} = z) \mathbb{P}^x(\tau_y^+ < \min(\tau_x^+, \sigma_{N-1})) \\
&= \tilde{p}(x, y) + \sum_{z \in S \setminus (T \cup \{x, y\})} \tilde{p}(x, z) \mathbb{P}^x(\tau_y^+ < \min(\tau_x^+, \sigma_{N-1}))
\end{aligned}$$

thus by iteration we obtain

$$\begin{aligned}
&= \sum_{k=1}^{N-1} \sum_{z_1, \dots, z_k \in S \setminus (T \cup \{x, y\})} \tilde{p}(x, z_1) \tilde{p}(z_1 \dots z_k) \tilde{p}(z_k, y) \\
&= \tilde{\mathbb{P}}^x(\tau_y^+ < \min(\tau_x^+, \sigma_N)) \xrightarrow{N \rightarrow \infty} \tilde{\mathbb{P}}^x(\tau_y^+ < \tau_x^+)
\end{aligned}$$

Thus $\mathbb{P}^x(\tau_y < \tau_x^+) = \tilde{\mathbb{P}}^x(\tau_y < \tau_x^+)$.

2. Let us first assume that p is irreducible, and so is \tilde{p} . Let μ and $\tilde{\mu}$ be their respective unique, positive stationary distributions. By Lemmas 6 and Lemma 7.1 we find

$$\frac{\mu(y)}{\mu(x)} = \frac{\mathbb{P}^x(\tau_y^+ < \tau_x^+)}{\mathbb{P}^y(\tau_x^+ < \tau_y^+)} = \frac{\tilde{\mathbb{P}}^x(\tau_y^+ < \tau_x^+)}{\tilde{\mathbb{P}}^y(\tau_x^+ < \tau_y^+)} = \frac{\tilde{\mu}(y)}{\tilde{\mu}(x)}$$

Summing this equation over $y \in \tilde{S}$ proves the irreducible case. To prove the general claim, let E_1, \dots, E_k be the essential classes of p , so the essential classes of \tilde{p} are the non empty sets among $E_1 \cap \tilde{S}, \dots, E_k \cap \tilde{S}$. For $i \leq k$ let μ_i ($\tilde{\mu}_i$) be the extension to S (\tilde{S}) of the unique stationary distribution of the irreducible $p|_{E_i \times E_i}$ ($\tilde{p}|_{E_i \cap \tilde{S} \times E_i \cap \tilde{S}}$). Let μ be a stationary distribution for p , it is well-known that μ is then a convex combination $\sum_{1 \leq i \leq k} \alpha_i \mu_i$, and it is straightforward to check that the convex combination $\tilde{\mu} := \sum_{1 \leq i \leq k} \frac{\sum_{y \in \tilde{S} \cap E_i} \mu(y)}{\sum_{y \in \tilde{S}} \mu(y)}$ witnesses the claim. Conversely, let $\tilde{\mu}$ be a stationary distribution of \tilde{p} , so it is a convex combination $\sum_{1 \leq i \leq k} \beta_i \tilde{\mu}_i$, and it is straightforward to check that the convex combination $\mu := \sum_{1 \leq i \leq k} \frac{L_i}{\sum_j L_j} \mu_i$ witnesses the claim, where $L_i := \beta_i \cdot \prod_{j \neq i} \frac{\mu_j(x_j)}{\tilde{\mu}_j(x_j)}$ for any $x_j \in \tilde{S} \cap E_j$. \square

Proof [Lemma 8] Let us first prove the claim for irreducible perturbations, and even in the following simpler case: let $x \in S$ be such that for all y and all $z \neq x$ we have $p(x, z) \cong \tilde{p}(x, z)$ and $p(z, y) = \tilde{p}(z, y)$; so $\mathbb{P}^z(\tau_y < \tau_x^+) = \tilde{\mathbb{P}}^z(\tau_y < \tau_x^+)$ for all $z \neq x$ since the paths leading from z to y without hitting x do not involve any step from x to another state. So

$$\begin{aligned}
\mathbb{P}^x(\tau_y^+ < \tau_x^+) &= \sum_{z \in S \setminus \{x\}} p(x, z) \mathbb{P}^z(\tau_y < \tau_x^+) \\
&\cong \sum_{z \in S \setminus \{x\}} \tilde{p}(x, z) \tilde{\mathbb{P}}^z(\tau_y < \tau_x^+) = \tilde{\mathbb{P}}^x(\tau_y^+ < \tau_x^+)
\end{aligned}$$

So by Lemmas 6, 47.2, and 47.3, and since the unique μ and $\tilde{\mu}$ are positive,

$$\frac{\mu(x)}{\mu(y)} = \frac{\mathbb{P}^y(\tau_x^+ < \tau_y^+)}{\mathbb{P}^x(\tau_y^+ < \tau_x^+)} \cong \frac{\tilde{\mathbb{P}}^y(\tau_x^+ < \tau_y^+)}{\tilde{\mathbb{P}}^x(\tau_y^+ < \tau_x^+)} = \frac{\tilde{\mu}(x)}{\tilde{\mu}(y)} \quad \text{for all } y \in S;$$

invoking this equation above twice shows that

$$\frac{\mu(z)}{\mu(y)} = \frac{\mu(z)}{\mu(x)} \cdot \frac{\mu(x)}{\mu(y)} \cong \frac{\tilde{\mu}(z)}{\tilde{\mu}(x)} \cdot \frac{\tilde{\mu}(x)}{\tilde{\mu}(y)} = \frac{\tilde{\mu}(z)}{\tilde{\mu}(y)} \quad \text{for all } z, y \in S;$$

and summing this second equation over $z \in S$ yields $\frac{1}{\mu(y)} \cong \frac{1}{\tilde{\mu}(y)}$, i.e., $\mu(y) \cong \tilde{\mu}(y)$ for all $y \in S$. The irreducible case is then

proved by induction on $n := |\{x \in S : \exists y \in S, x \neq y \wedge p(x, y) \neq \tilde{p}(x, y)\}|$, which trivially holds for $n = 0$. For $0 < n$ let distinct $x, y \in S$ be such that $p(x, y) \neq \tilde{p}(x, y)$, and for all $z \in S \times S \setminus \{x, y\}$ let $\hat{p}(z, y) := p(z, y)$, and $\hat{p}(x, y) := \tilde{p}(x, y)$. By the simpler case $\mu \cong \hat{\mu}$; by induction hypothesis $\hat{\mu} \cong \tilde{\mu}$; so $\mu \cong \tilde{\mu}$ by transitivity of \cong .

Let us now prove the general claim by induction on the number of the non-zero transition maps of p that have zeros. Base case, all the non-zero maps are positive. Let $E'_1, \dots, E'_{k'}$ be the sink SCCs of the graph on S with arc (x, y) if $p(x, y)$ is non-zero. For $i \leq k'$ let μ_i ($\tilde{\mu}_i$) be the unique stationary distribution map of the irreducible $p|_{E'_i \times E'_i}$ ($\tilde{p}|_{E'_i \times E'_i}$). Since $p(x, y)|_{E'_i \times E'_i} \cong \tilde{p}(x, y)|_{E'_i \times E'_i}$ for all $x \neq y$, the irreducible case implies $\mu_i \cong \tilde{\mu}_i$. Clearly μ is a convex combination of the μ_i , and the convex combination $\tilde{\mu}$ of the $\tilde{\mu}_i$ with the same coefficients is a stationary distribution map for \tilde{p} , and $\mu \cong \tilde{\mu}$ by Lemma 47.3.

Inductive case. Let $p(z, t)$ be a non-zero function with support $J \subsetneq I$. Up to focusing we may assume that 0 is a limit point of both J and $I \setminus J$. By induction hypothesis on $p|_J \cong \tilde{p}|_J$ and $p|_{I \setminus J} \cong \tilde{p}|_{I \setminus J}$ we obtain two distribution maps μ_J and $\tilde{\mu}_{I \setminus J}$ that can be combined to witness the claim. \square

Proof [Observation 10] Let $p(x, y)$ be in the essential graph. By the definition of \precsim and finiteness of the state space, let positive b_{xy} and ϵ_{xy} be such that $p(x, z) \leq b_{xy} \cdot p(x, y)$ for all $\epsilon < \epsilon_{xy}$ and $z \in S$. So for all $\epsilon < \epsilon_{xy}$ we have $1 = \sum_{z \in S} p_\epsilon(x, z) \leq |S| \cdot b_{xy} \cdot p_\epsilon(x, y)$. Now let $b(\epsilon_0)$ be the maximum (minimum) of the b_{xy} (ϵ_{xy}) for (x, y) in the essential graph. Thus $0 < (b \cdot |S|)^{-|S|} \leq p_\epsilon(\gamma)$ for all $\epsilon < \epsilon_0$. \square

Proof [Lemma 12] For all $y \in T$ let x_y be an essential state reachable from y in the essential graph. So $c \cdot \mu(y) \leq \mu(x_y)$ for all $y \in T$ by Lemma 6. Therefore $1 \leq c^{-1} \sum_{y \in T} \mu(x_y) + \sum_{x \in S \setminus T} \mu(x)$, and the claim is proved by further approximation. \square

Lemma 48 below is a technical tool proved by a standard argument in Markov chain theory.

Lemma 48 Let a perturbation with state space S satisfy Assumption 11, and let x be in some essential class E . Then for all $n \in \mathbb{N}$

$$\mathbb{P}^x(\tau_{(S \setminus E) \cup \{x\}}^+ > n) \leq (1 - c)^{\lfloor \frac{n}{|S|} \rfloor}$$

Proof Let $\tau^* := \tau_{(S \setminus E) \cup \{x\}}^+$. For every $y \in E$ let $\gamma_y \in \Gamma_E(y, x)$ be in the essential graph. $c < p(\gamma_y)$ by Assumption 11, so $\max_{y \in E} \mathbb{P}_\epsilon^y(\tau^* > |S|) \leq 1 - c < 1$, so for all $k \in \mathbb{N}$

$$\mathbb{P}^x(\tau^* > k|S|) \leq (\max_{y \in E} \mathbb{P}_\epsilon^y(\tau^* > |S|))^k$$

$\leq (1 - c)^k$ by the strong Markov property, so

$$\mathbb{P}^x(\tau^* > n) \leq \mathbb{P}^x(\tau^* > |S| \cdot \lfloor \frac{n}{|S|} \rfloor) \leq (1 - c)^{\lfloor \frac{n}{|S|} \rfloor} \text{ for all } n \in \mathbb{N}.$$

\square

Proof [Lemma 15] Up to focusing let p satisfy Assumption 11. The second statement boils down to Lemma 47.3. For the first statement, for every $z \in E$ let $\gamma_z \in \Gamma_E(x, z)$ be a path in the essential graph, so $c < p(\gamma_z)$ by Assumption 11. Let $\tau^* := \tau_{(S \setminus E) \cup \{x\}}^+$, so

$$\begin{aligned}
c \cdot p(z, y) &\leq c \cdot \mathbb{P}^z(X_{\tau^*} = y) \leq p(\gamma_z) \cdot \mathbb{P}^z(X_{\tau^*} = y) \\
&\leq \mathbb{P}^x(X_{\tau^*} = y) = \tilde{p}(\cup E, y)
\end{aligned}$$

which implies $c \cdot \max_{z \in E} p(z, y) \leq \tilde{p}(\cup E, y)$ i.e. half of the statement $\tilde{p}(\cup E, y) \cong \max_{z \in E} p(z, y)$. Now, for every positive

natural N let $\mathbf{A}_N := \{x\} \times (E \setminus \{x\})^{N-1} \times \{y\}$.

$$\begin{aligned}\mathbb{P}^x(X_{\tau^*} = y, \tau^* = N) &= \sum_{\gamma \in \mathbf{A}_N} p(\gamma) \\ &\leq \max_{z \in E} p(z, y) \sum_{\gamma \in \mathbf{A}_N} p(\gamma) \\ &= \max_{z \in E} p(z, y) \mathbb{P}^x(\tau^* \geq N - 1)\end{aligned}$$

Let $q := (1 - c)^{|S|} < 1$, so $\mathbb{P}^x(\tau^* \geq N) \leq (1 - c)^{-1} \cdot q^N$ by Lemma 48, and

$$\begin{aligned}\mathbb{P}^x(X_{\tau^*} = y) &= \sum_{N=1}^{\infty} \mathbb{P}^x(X_{\tau^*} = y, \tau^* = N) \\ &\leq \max_{z \in E} p(z, y) \sum_{N=0}^{\infty} (1 - c^{-1}) \cdot q^N \\ &= (1 - c)^{-1} (1 - q)^{-1} \cdot \max_{z \in E} p(z, y)\end{aligned}$$

□

Proof [Proposition 16] Up to focusing let p satisfy Assumption 11.

1. Let y be in the set of the transient states T , so there exists $x \notin T$ and $\gamma \in \Gamma_T(y, x)$ such that γ is also in the essential graph. By Assumption 11 this implies

$$0 < \liminf_{\epsilon \rightarrow 0} p_\epsilon(\gamma) \leq \liminf_{\epsilon \rightarrow 0} \mathbb{P}_\epsilon^y(\tau_x^+ < \tau_y^+).$$

On the other hand, let E be the essential class of x . Lemma 15 implies that

$$\begin{aligned}\mathbb{P}^x(\tau_y^+ < \tau_x^+) &\leq 1 - \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}^+} = x) \\ &= \sum_{z \notin E} \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}^+} = z) \cong \sum_{z \notin E} \max_{t \in E} p(t, z)\end{aligned}$$

Since E is an essential class, for all $t \in E$ and $z \notin E$ the function $p(t, z)$ is not \precsim -maximal. So $\liminf_{\epsilon \rightarrow 0} p_\epsilon(t, z) = 0$ by definition of the essential graph, and $\liminf_{\epsilon \rightarrow 0} \sum_{z \notin E} \max_{t \in E} p_\epsilon(t, z) = 0$ by Assumption 5, Lemma 47.6, and finiteness. By combining this with the two inequalities above one obtains

$$\liminf_{\epsilon \rightarrow 0} \frac{\mathbb{P}_\epsilon^x(\tau_y^+ < \tau_x^+)}{\mathbb{P}_\epsilon^y(\tau_x^+ < \tau_y^+)} = 0$$

and noting the following by Lemma 6 allows us to conclude.

$$\tilde{\mu}_\epsilon(y) \leq \frac{\tilde{\mu}_\epsilon(y)}{\tilde{\mu}_\epsilon(x)} = \frac{\mathbb{P}_\epsilon^x(\tau_y^+ < \tau_x^+)}{\mathbb{P}_\epsilon^y(\tau_x^+ < \tau_y^+)}$$

2. By Assumption 11 (twice) and Lemma 6.

□

Proof [Lemma 17] Let us first prove the first conjunct. On the one hand $\mathbb{P}^y(\tau_x < \tau_y) \leq \mathbb{P}^y(\tau_E < \tau_y) = \tilde{\mathbb{P}}^y(\tau_{\tilde{x}} < \tau_y)$, and on the other hand $\mathbb{P}^y(\tau_x < \tau_y) \geq \mathbb{P}^y(\tau_E < \tau_y) \cdot \min_{z \in E} \mathbb{P}^z(\tau_x < \tau_y)$ by the strong Markov property. Since $\lim_{\epsilon \rightarrow 0} \mathbb{P}^z(\tau_x < \tau_y) = 1$ for all z in the essential class E , $\mathbb{P}^y(\tau_{\tilde{x}} < \tau_y)$ and $\mathbb{P}^y(\tau_x < \tau_y)$ are even asymptotically equivalent.

Let us now prove the second conjunct. Let $\tau_{E,0}^* := 0$ and $\tau_{E,n+1}^* := \inf\{t > \tau_{E,n}^* \mid X_t \in E \wedge \exists s \in]\tau_{E,n}^*, t[: X_s \notin E\}$ for all $n \in \mathbb{N}$. Informally, when starting in E , $\tau_{E,n}^*$ is the first time that the chain is in E after n stays outside that are separated by

stays inside.

$$\begin{aligned}\tilde{\mathbb{P}}^{\cup E}(\tau_y < \tau_{\cup E}) &= \mathbb{P}^x(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*) \\ &\quad \text{by definition of the essential collapse, and } \tau_{E,1}^* \\ &\leq \mathbb{P}^x(\tau_y < \tau_x), \text{ showing half of the first conjunct.}\end{aligned}$$

For the other half, let X satisfy Assumption 11 up to focusing. For all $z \in E \setminus \{x\}$ there is a simple path in the essential graph from x to z , so $c < \mathbb{P}^x(\tau_z < \tau_{S \setminus E \cup \{x\}})$, and by the strong Markov property

$$\begin{aligned}\mathbb{P}^x(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*) &\geq \mathbb{P}^x(\tau_z < \tau_{S \setminus E \cup \{x\}}) \cdot \mathbb{P}^z(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*) \\ &\geq c \cdot \mathbb{P}^z(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*)\end{aligned}\tag{1}$$

For all $z \in E \setminus \{x\}$ there is also a simple path in the essential graph from z to x , so $\mathbb{P}^z(\tau_{S \setminus E} < \tau_x) \leq 1 - c$. Also note that $\mathbb{P}^z(\tau_{E,1}^* < \tau_x) \leq \mathbb{P}^z(\tau_{S \setminus E} < \tau_x)$, and let us show by induction on n that $\mathbb{P}^z(\tau_{E,n}^* < \tau_x) \leq (1 - c)^n$, which holds for $n = 0$.

$$\begin{aligned}\mathbb{P}^x(\tau_{E,n+1}^* < \tau_x) &\leq \mathbb{P}^x(\tau_{E,n}^* < \tau_x) \cdot \max_{z \in E} \mathbb{P}^z(\tau_{E,1}^* < \tau_x)\end{aligned}\tag{2}$$

by the strong Markov property.

$$\begin{aligned}&\leq \mathbb{P}^x(\tau_{E,n}^* < \tau_x) \cdot (1 - c) \text{ by the remark above.} \\ &\leq (1 - c)^{n+1} \text{ by induction hypothesis.}\end{aligned}\tag{3}$$

Let us now conclude about the second half of the first conjunct.

$$\mathbb{P}^x(\tau_y < \tau_x) = \sum_{n=0}^{\infty} \mathbb{P}^x(\tau_y < \tau_x, \tau_{E,n}^* < \tau_y < \tau_{E,n+1}^*)$$

by a case disjunction.

$$\leq \sum_{n=0}^{\infty} \mathbb{P}^x(\tau_y < \tau_x, \tau_{E,n}^* < \tau_x, \tau_y < \tau_{E,n+1}^*)$$

since the new conditions are weaker.

$$\leq \sum_{n=0}^{\infty} \mathbb{P}^x(\tau_{E,n}^* < \tau_x) \max_{z \in E} \mathbb{P}^z(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*)$$

by the strong Markov property.

$$\leq c^{-1} \mathbb{P}^x(\tau_y < \tau_x, \tau_y < \tau_{E,1}^*) \cdot \sum_{n=0}^{\infty} (1 - c)^n$$

by inequalities 1 and 3.

$$\leq c^{-2} \tilde{\mathbb{P}}^{\cup E}(\tau_y < \tau_{\cup E})$$

□

Proof [Proposition 18] Up to focusing let p satisfy Assumption 11.

1. Let us first assume that p is irreducible, and so is \tilde{p} by Observation 14. So both p and \tilde{p} have unique, positive stationary distribution maps. Let us prove that $\tilde{\mu}(\tilde{x}) \cong \mu(x)$ where $\tilde{x} := \cup E$ and $\tilde{\mu}(y) \cong \mu(y)$ for all $y \in S \setminus E$. For $y \in S \setminus E$, Lemma 6 and Lemma 17 imply the following.

$$\frac{\mu(y)}{\mu(x)} = \frac{\mathbb{P}^x(\tau_y^+ < \tau_x^+)}{\mathbb{P}^y(\tau_x^+ < \tau_y^+)} \cong \frac{\tilde{\mathbb{P}}^{\tilde{x}}(\tau_y^+ < \tau_{\tilde{x}}^+)}{\mathbb{P}^y(\tau_{\tilde{x}}^+ < \tau_y^+)} = \frac{\tilde{\mu}(y)}{\tilde{\mu}(\tilde{x})}\tag{4}$$

Summing the above equation over $y \in S \setminus E$ and adding $\frac{\mu(x)}{\mu(x)} = \frac{\tilde{\mu}(\tilde{x})}{\tilde{\mu}(\tilde{x})}$ yields $(1 - \bar{\mu})\mu(x)^{-1} \cong \tilde{\mu}(x)^{-1}$, where $\bar{\mu} := \sum_{z \in S \setminus E} \mu(z)$. So by Definition 3, let a and b be positive real numbers such that $a \cdot \tilde{\mu}(\tilde{x})^{-1} \leq (1 - \bar{\mu})\mu(x)^{-1} \leq b \cdot \tilde{\mu}(\tilde{x})^{-1}$ on a neighborhood of 0, which yields $a \cdot \mu(x) \leq \tilde{\mu}(\tilde{x}) \leq b \cdot \mu(x) + \bar{\mu}$. Since $\bar{\mu} \leq b' \cdot \mu(x)$ for some b' by Proposition 16.2, $\tilde{\mu}(\tilde{x}) \cong \mu(x)$. Now by Lemmas 47.2 and 47.3, let us rewrite $\mu(x)$ with $\tilde{\mu}(\tilde{x})$ in Equation 4, which shows the claim for irreducible perturbations.

Let us now prove the general claim by induction on the number of the non-zero transition maps of p that have zeros. Base case, all the non-zero maps are positive. Let $E'_1 \dots, E'_{k'}$ be the sink SCCs of the graph on S with arc (x, y) if $p(x, y)$ is non-zero. The essential graph is included in this digraph, and the essential class E is included in E'_j for some $j \in \{1, \dots, k'\}$. Let $\tilde{E}'_j := \{E\} \cup E'_j \setminus E$ and for all $i \neq j$ let $\tilde{E}'_i := E'_i$. For all $i \in \{1, \dots, k'\}$ let $\mu_{E'_i}$ ($\tilde{\mu}_{\tilde{E}'_i}$) be the extension to S ($\{E\} \cup S \setminus E$), by the zero-function outside of E'_i (\tilde{E}'_i), of the unique stationary distribution of $p|_{E'_i \times E'_i}$ ($\tilde{p}|_{\tilde{E}'_i \times \tilde{E}'_i}$), and by the irreducible case above let $\tilde{\mu}_{\tilde{E}'_j}$ ($\mu_{E'_j}$) be the corresponding unique distribution of $\tilde{p}|_{\{E\} \cup E'_j \setminus E}$ after ($p|_{E'_j \times E'_j}$ before) the essential collapse. Since μ ($\tilde{\mu}$) is a convex combination $\sum_{1 \leq i \leq k'} \alpha_i \mu_{E'_i}$ ($\sum_{1 \leq i \leq k'} \alpha_i \tilde{\mu}_{\tilde{E}'_i}$), it is easy to check that $\tilde{\mu} := \sum_{1 \leq i \leq k'} \alpha_i \tilde{\mu}_{\tilde{E}'_i}$ ($\mu := \sum_{1 \leq i \leq k'} \alpha_i \mu_{E'_i}$) is a witness for the base case.

Inductive case. Let $p(z, t)$ be a non-zero function with support $J \subsetneq I$. Up to focusing we may assume that 0 is a limit point of both J and $I \setminus J$. By induction hypothesis on $p|_J \cong \tilde{p}|_J$ and $p|_{I \setminus J} \cong \tilde{p}|_{I \setminus J}$ we obtain two distribution maps $\tilde{\mu}_J$ and $\tilde{\mu}_{I \setminus J}$ (μ_J and $\mu_{I \setminus J}$) that can be combined to witness the claim.

2. By Propositions 16.1 and 18.1 in both cases, and also by Proposition 16.2 if $y \in E$.

□

Proof [Proposition 21] Up to focusing let p satisfy Assumption 11. By induction on the number of the non-zero transition maps of p that have zeros. Base case, all the non-zero maps are positive. Let $E'_1 \dots, E'_{k'}$ be the sink SCCs of the graph on S with arc (x, y) if $p(x, y)$ is non-zero. Note that this digraph includes the essential graph, and that $\delta(p|_{E'_i \times E'_i}) = \delta(p)|_{E'_i \cap S \setminus T \times E'_i \cap S \setminus T}$. Moreover, by Lemmas 7.2 the stable states of each $p|_{E'_i \times E'_i}$ are also stable for $\delta(p|_{E'_i \times E'_i})$, and the converse holds by Lemmas 7.2 and 12. Therefore a state is stable for p iff it is stable for some $p|_{E'_i \times E'_i}$ iff it is stable for some $\delta(p)|_{E'_i \cap S \setminus T \times E'_i \cap S \setminus T}$ iff it is stable for $\delta(p)$.

Inductive case. Let $p(z, t)$ be a non-zero function with support $J \subsetneq I$. Up to focusing we may assume that 0 is a limit point of both J and $I \setminus J$. By induction hypothesis $p|_J$ and $\delta(p)|_J$ have the same stable states, and likewise for $p|_{I \setminus J}$ and $\delta(p)|_{I \setminus J}$, which shows the claim. □

Lemma 49 below is a generalization to the reducible case of Proposition 2.8 from [1].

Lemma 49 Let A be a finite subset of the state space S of a Markov chain. Then for all $x \in A$ and $y \in S \setminus A$

$$\mathbb{P}^x(X_{\tau_{S \setminus A}} = y) = \sum_{\gamma \in \Gamma_A(x, y), p(\gamma) > 0} \prod_{i=1}^{|\gamma|-1} \frac{p(\gamma_i, \gamma_{i+1})}{1 - \mathbb{P}^{\gamma_i}(X_{\tau_{(S \setminus A) \cup \{\gamma_1, \dots, \gamma_i\}}} = \gamma_i)}$$

Proof We proceed by induction on $|A|$. The claim trivially holds for $|A| = 0$; so now let $x \in A$. The strong Markov property gives

$$\begin{aligned} \mathbb{P}^x(X_{\tau_{S \setminus A}} = y) &= \mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = y) \\ &\quad + \mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = x) \cdot \mathbb{P}^x(X_{\tau_{S \setminus A}} = y) \end{aligned}$$

If $p(\gamma) = 0$ for all $\gamma \in \Gamma_A(x, y)$, the claim boils down to $0 = 0$, so let us assume that there exists $\gamma \in \Gamma_A(x, y)$ with $p(\gamma) = 0$, so $\mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = x) < 1$, and the above equation may be rearranged into

$$\mathbb{P}^x(X_{\tau_{S \setminus A}} = y) = \frac{\mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = y)}{1 - \mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = x)}$$

where the numerator may be decomposed as $p(x, y) + \sum_{z \in A \setminus \{x\}} p(x, z) \mathbb{P}^z(X_{\tau_{(S \setminus A) \cup \{x\}}} = y)$. By the induction hypothesis for the set $A \setminus \{x\}$ let us rewrite $\mathbb{P}^z(X_{\tau_{(S \setminus A) \cup \{x\}}} = y)$ for all $z \in A \setminus \{x\}$, and obtain the equation below that may be re-indexed to yield the claim.

$$\begin{aligned} \mathbb{P}^x(X_{\tau_{S \setminus A}} = y) &= \frac{p(x, y)}{1 - \mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = x)} + \sum_{z \in A \setminus \{x\}, p(x, z) > 0} \\ &\quad \sum_{\gamma \in \Gamma_{A \setminus \{x\}}(z, y), p(\gamma) > 0} \frac{p(x, z) \cdot \Pi}{1 - \mathbb{P}^x(X_{\tau_{(S \setminus A) \cup \{x\}}} = x)} \\ &\quad \text{where } \Pi := \prod_{i=1}^{|\gamma|-1} \frac{p(\gamma_i, \gamma_{i+1})}{1 - \mathbb{P}^{\gamma_i}(X_{\tau_{(S \setminus A) \cup \{x, \gamma_1, \dots, \gamma_i\}}} = \gamma_i)} \end{aligned}$$

□

Proof [Lemma 22] Up to focusing let p satisfy Assumption 11. Let $\tau^* := \tau_{S \setminus T}^+$, and consider

$$\mathbb{P}^x(X_{\tau^*} = y) = p(x, y) + \sum_{z \in T} p(x, z) \mathbb{P}^z(X_{\tau^*} = y).$$

For all $z \in T$ there are $z' \in S \setminus T$ and $\gamma_z \in \Gamma_T(z, z')$ in the essential graph, so for all $K \subseteq S$ we have $\mathbb{P}^z(X_{\tau_{(S \setminus T) \cup K}^+} = z) \leq \mathbb{P}^z(X_{\tau_{(S \setminus T) \cup \{z\}}} = z) \leq 1 - p(\gamma) < 1 - c$. So by Lemma 49

$$\mathbb{P}^z(X_{\tau^*} = y) \leq \sum_{\gamma \in \Gamma_T(z, y)} p(\gamma) \cdot c^{-|\gamma|} \leq c^{-|T|} \sum_{\gamma \in \Gamma_T(z, y)} p(\gamma).$$

Since $\mathbb{P}^z(X_{\tau^*} = y) \geq \sum_{\gamma \in \Gamma_T(z, y)} p(\gamma)$, thus $\mathbb{P}^z(X_{\tau^*} = y) \cong \sum_{\gamma \in \Gamma_T(z, y)} p(\gamma)$, an by Lemma 47.5 we can replace the sum with the maximum. □

Proof [Observation 24] Let $\epsilon \in I$. If $g(\epsilon) \neq 0$ then $((f \div_n g) \cdot g)(\epsilon) = \frac{f(\epsilon)}{g(\epsilon)} \cdot g(\epsilon) = f(\epsilon)$; if $g(\epsilon) = 0$ then $f(\epsilon) = 0 = (f \div_n g)(\epsilon) \cdot g(\epsilon)$. □

Proof [Proposition 26] Let m be as in Definition 25 and let $J \subseteq I$ be its support.

1. $\sigma(p)(x, y)|_{I \setminus J} = |S|^{-1}$ for all $x, y \in S$, so $\sigma(p)_\epsilon$ is stochastic for all $\epsilon \in I \setminus J$. Let $\epsilon \in J$. On the one hand $\sum_{y \in S} \sigma(p)_\epsilon(x, y) = \frac{p_\epsilon(x, x) + m - 1}{m} + \sum_{y \in S \setminus \{x\}} \frac{p_\epsilon(x, y)}{m} = 1$, on the other hand $\sum_{y \in S \setminus \{x\}} \sigma(p)_\epsilon(x, y) = \sum_{y \in S \setminus \{x\}} \frac{p_\epsilon(x, y)}{|S| \cdot \max\{p_\epsilon(t, z) | z, t \in S \wedge z \neq t\}} \leq \frac{(|S|-1)}{|S|} < 1$, so $\sigma(p)_\epsilon$ is stochastic. If 0 is not a limit point of J , it is clear that $\sigma(p)(x, y) \cong 1$ for all $x, y \in S$, and $\sigma(p)$ satisfies Assumption 5. If 0 is a limit point of J , then $p(x, y)|_J \lesssim p(x', y')|_J$ implies $\sigma(p)(x, y)|_J \lesssim \sigma(p)(x', y')|_J$ (since m is non-zero on J), so $\sigma(p)$ satisfies Assumption 5 by Lemma 47.7 (since $p|_J$ and then $\sigma(p)|_J$ do).

2. Let us first prove the claim if $J = I$. The equation below shows that $\mu \cdot p = \mu$ iff $\mu \cdot \sigma(p) = \mu$, for all distribution maps μ on $S = \{x_1, \dots, x_n\}$, so p and $\sigma(p)$ have the same stable states.

$$\begin{aligned} (\mu \cdot \sigma(p))_j &= \frac{p(x_j, x_j) + m - 1}{m} \cdot \mu_j + \sum_{i \neq j} \frac{p(x_i, x_j) \mu_i}{m} \\ &= \frac{(m-1)\mu_j + \sum_i p(x_i, x_j) \mu_i}{m} \end{aligned}$$

Let us now prove the claim if $J \subsetneq I$. The case where 0 is not a limit point of $I \setminus J$ amounts, up to focusing, to the case $J = I$, so let us assume that 0 is a limit point of $I \setminus J$. For all $\epsilon \in I \setminus J$ we have $p_\epsilon(x, y) = 0$ for all states $x \neq y$, so all distributions are stationary for p_ϵ . Moreover, the uniform

distribution is stationary for $\sigma(p)_\epsilon$, so all states are stable for $p|_{I \setminus J}$ and $\sigma(p)|_{I \setminus J}$. Therefore, if 0 is not a limit point of J , we are done, and if it is, we are also done by Lemma 47.8.

3. Let distinct $x, y \in S$ be such that $p(z, t) \lesssim p(x, y)$ for all distinct $z, t \in S$. So $p(x, y) \cong \max\{p(z, t) \mid z, t \in S \wedge z \neq t\}$ by Lemma 47.6, so $\sigma(p)(x, y) \cong p(x, y) \div_{|S|} |S| \cdot p(x, y) = \frac{1}{|S|} \cong 1$, so (x, y) is in the essential graph of $\sigma(p)$.

□

Lemma 50 *Let a perturbation p with state space S satisfy Assumption 5, let E_1, \dots, E_k be its essential classes, and for all i let $x_i \in E_i$. The state x is stable for p iff x belongs to some E_i such that $\cup E_i$ is stable for $\delta \circ \kappa(\dots \kappa(\kappa(\sigma(p), x_1), x_2) \dots, x_k)$.*

Proof [Theorem 27] By applying Lemma 50 recursively. If p is the identity matrix then all states are stable. Otherwise the essential graph of $\sigma(p)$ is non-empty, so either one essential class is not a singleton, or one state is transient. If there is a non-singleton essential class, the corresponding essential collapse decreases the number of states; if one state is transient, the transient deletion decreases the number of states. Since these transformations do not increase the number of states, $\delta \circ \kappa(\dots \kappa(\kappa(\sigma(p), x_1), x_2) \dots, x_k)$ has fewer states than p , whence termination of the recursion on an identity perturbation whose non-empty state space corresponds to the stable states of p . □

Proof [Observation 31] If $f \leq 0$ then $f = (f \div 0) \cdot 0 = 0$. Also, $(f \div 1) = (f \div 1) \cdot 1 = f$. □

Proof [Lemma 32]

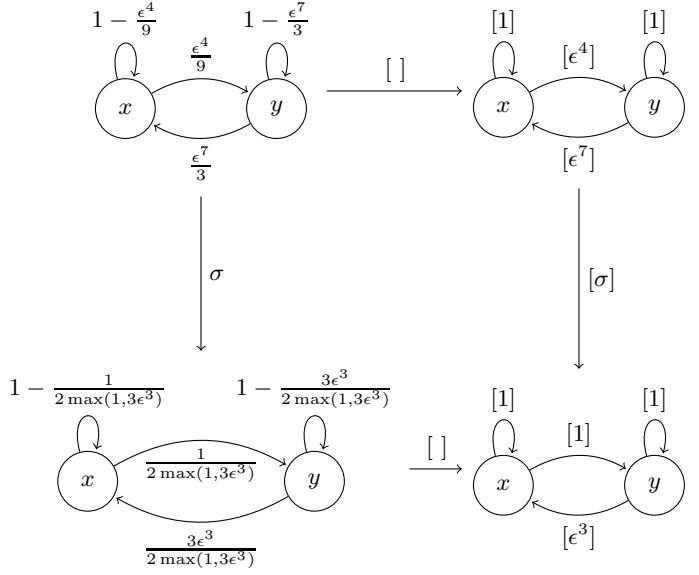
1. By Lemma 47.7 with J the support of g and $I \setminus J$, then by Lemmas 47.2 and 47.3.
2. By Observations 29.2 and 24, and since $[\epsilon \mapsto 1]$ is the $[\lesssim]$ -maximum of $[G \cup \{\epsilon \mapsto 0\}]$.

□

Proof [Lemma 35]

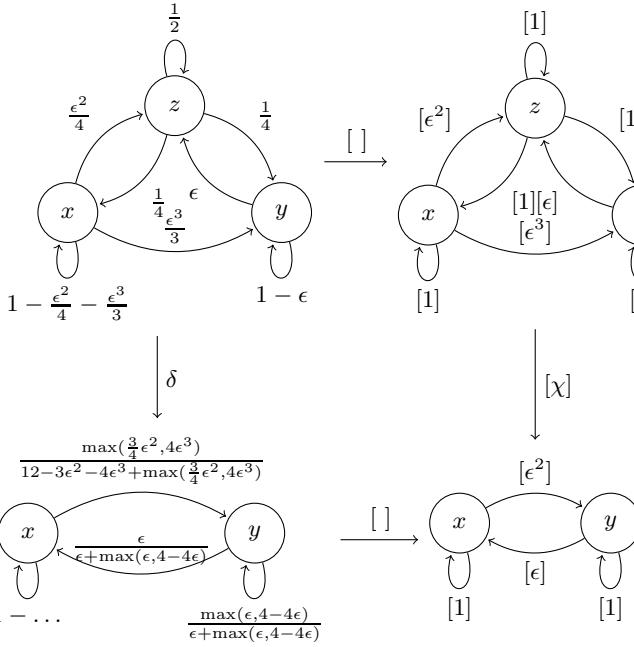
1. Clear by comparing Definitions 9 and 33.1.
2. Let $[p](z, t) = \max\{[p](z', t') : (z', t') \in S \times S \wedge z' \neq t'\}$. If $x \neq y$ then

$$\begin{aligned} [\sigma]([p])(x, y) &= [p](x, y)[\div][p](z, t) \text{ by definition of } [\sigma], \\ &= [p(x, y)][\div][p(z, t)] \text{ by definition of } [p], \\ &= [p(x, y) \div_{|S|} p(z, t)] \text{ by Lemma 32,} \\ &= [p(x, y) \div_{|S|} \max\{p(z, t) \mid z, t \in S \wedge z \neq t\}] \text{ by Lemma 32 and Lemma 47.6,} \\ &= [\sigma(p)](x, y) \text{ by definition of } \sigma. \end{aligned}$$

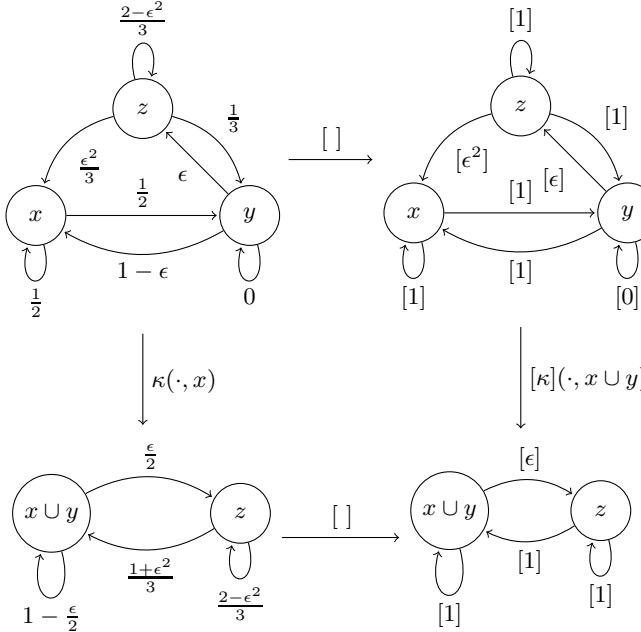


3. The essential classes of p are singletons since $\delta(p)$ is well-defined. Let $\{x\}$ and $\{y\}$ be distinct essential classes of p , and of $[p]$ by Lemma 35.1. Let $M := \max_{[\lesssim]} \{[p](\gamma) : \gamma \in \Gamma_T(x, y)\}$, so $M = [\max\{p(\gamma) : \gamma \in \Gamma_T(x, y)\}]$ by Lemma 47.6. Note that $p(x, x) \cong 1$ since $\{x\}$ is an essential class of p , so $\sum_{z \in S \setminus T} \max\{p(\gamma) : \gamma \in \Gamma_T(x, z)\} \cong 1$ too. So

$$\begin{aligned} [\chi]([p])(\{x\}, \{y\}) &= M \text{ by definition of } [\chi], \\ &= M[\div][1] \text{ by Observation 31,} \\ &= M[\div][\sum_{z \in S \setminus T} \max\{p(\gamma) : \gamma \in \Gamma_T(x, z)\}] \text{ by a remark above,} \\ &= [\max\{p(\gamma) : \gamma \in \Gamma_T(x, y)\}][\div] \\ &\quad [\sum_{z \in S \setminus T} \max\{p(\gamma) : \gamma \in \Gamma_T(x, z)\}] \text{ by a remark above,} \\ &= [\max\{p(\gamma) : \gamma \in \Gamma_T(x, y)\} \div_{|S|} \\ &\quad \sum_{z \in S \setminus T} \max\{p(\gamma) : \gamma \in \Gamma_T(x, z)\}] \text{ by Lemma 32,} \\ &= [\delta(p)](x, y) \text{ by definition of } \delta. \end{aligned}$$



4. Let $x, y \in S \setminus E_1$. First, $[\kappa]([p], E_i)(x, y) = [p](x, y) = [p(x, y)] = [\kappa(p, x_i)(x, y)]$ by definitions of $[\kappa]$, $[p]$, and κ . Also $[\kappa]([p], E_i)(\cup E_i, y) = \max_{[z]} \{[p](x, y) \mid x \in E_i\} = [\max\{p(x, y) \mid x \in E_i\}] = [\kappa(p, x_i)(\cup E_i, y)]$ by definition, Lemma 47.6, and Lemma 15. Likewise $[\kappa]([p], E_i)(y, \cup E_i) = \max_{[z]} \{[p](y, x) \mid x \in E_i\} = [\max\{p(y, x) \mid x \in E_i\}] = [\kappa(p, x_i)(y, \cup E_i)]$.



5. Let $P := [p]$ and $\leq := [\preceq]$, and let us prove the claim abstractly. First note that $[\chi] \circ [\kappa](P, E_1)$ and $[\chi](P)$ have the same state space $\{\cup E_1, \dots, \cup E_k\}$. For $i, j \neq 1$ the definition of $[\chi]$ gives $[\chi] \circ [\kappa](P, E_1)(\cup E_j, \cup E_j) = \max_{\gamma \in \Gamma_T(E_i, E_j)} \{[\kappa](P, E_1)(\gamma) : \gamma \in \Gamma_T(E_i, E_j)\}$, where $T := S \setminus \cup_i E_i$. It is equal to $[\chi](P)(\cup E_j, \cup E_j)$ since $[\kappa](P, E_1)(x, y) = P(x, y)$ for

all $x, y \in S \setminus E_1$, which then also holds for paths $\gamma \in \Gamma_T(E_i, E_j)$. Let us now show that $[\chi](P)(\cup E_1, \cup E_j) = [\chi] \circ [\kappa](P, E_1)(\cup E_1, \cup E_j)$. On the one hand for all paths $x\gamma \in \Gamma_T(E_i, E_j)$ we have $P(x\gamma) \leq [\kappa](P, E_1)((\cup E_1)\gamma)$ since $[\kappa](P, E_1)(\cup E_1, y) := \max_{\gamma \in T^* \times E_j} \{P(x, y) : x \in E_1\}$, and on the other hand for every $\gamma \in T^* \times E_j$ there exists $x \in E_1$ such that $[\kappa](P, E_1)((\cup E_1)\gamma) = P(x\gamma)$. So

$$\begin{aligned} [\chi](P)(\cup E_1, \cup E_j) &= \max_{\gamma \in \Gamma_T(E_1, E_j)} \{P(\gamma) : \gamma \in \Gamma_T(E_1, E_j)\} \text{ by definition,} \\ &= \max_{\gamma \in \Gamma_T(E_1, E_j)} \{[\kappa](P, E_1)(\gamma) : \gamma \in \Gamma_T(E_1, E_j)\} \\ &\quad \text{by the remark above,} \\ &= [\chi] \circ [\kappa](P, E_1)(\cup E_1, \cup E_j) \text{ by definition.} \end{aligned}$$

The equality $[\chi](P)(\cup E_i, \cup E_1) = [\chi] \circ [\kappa](P, E_1)(\cup E_i, \cup E_1)$ can be proved likewise.

6. Let us first prove $[\delta \circ \kappa(\dots \kappa(\kappa(p, x_1), x_2) \dots, x_k)](\cup E_i, \cup E_j) = [\chi]([p])(\cup E_i, \cup E_j)$ for all $i \neq j$ by induction on the number k' of non-singleton essential classes. Since collapsing a singleton class has no effect, the claim holds for $k' = 0$ by Lemma 35.3, so let us assume that it holds for some arbitrary k' and that p has $k' + 1$ non-singleton essential classes. One may assume up to commuting and renaming that E_1 is not a singleton. Since $\kappa(\kappa(p, x_1), \cup E_1) = \kappa(p, x_1)$, also $\delta \circ \kappa(\dots \kappa(\kappa(p, x_1), x_2) \dots, x_k) = \delta \circ \kappa(\dots \kappa(\kappa(p, x_1), \cup E_1), x_2) \dots, x_k)$. So $[\delta \circ \kappa(\dots \kappa(\kappa(p, x_1), x_2) \dots, x_k)](\cup E_i, \cup E_j) = [\chi]([\kappa(p, x_1)])(\cup E_i, \cup E_j)$ for all $i \neq j$ by induction hypothesis. Moreover, $[\chi](\kappa(p, x_1)) = [\chi](\kappa([p], E_1)) = [\chi]([p])$ by Lemmas 35.4 and 35.5. Therefore $[\chi](\sigma(p))(\cup E_i, \cup E_j) = [\delta \circ \kappa(\dots \kappa(\kappa(\sigma(p), x_1), x_2) \dots, x_k)](\cup E_i, \cup E_j)$ for all $i \neq j$ by Lemma 35.2 and Observation 34. \square

Proof [Proposition 37] Line 2 from Algorithm 1 is performed once and takes n steps; Line 3 takes one step and is performed n^2 times. Let us now focus on the recursive function HubRec. If all the arcs of the input are labelled with 0, the algorithm terminates; if not, $p(s, t) = 1$ at least for some distinct $s, t \in S$ after the outgoing scaling at Line 9, so either the strongly connected component of s is not a sink, or s is in the same strongly connected component as t , which implies in both cases that there are fewer $\cup S_i$ than vertices in S , and subsequently that HubRec is recursively called at most n times for an input with n vertices. Lines 7, 9, 10, 12, 13, 14, and 15 take at most $O(n^2)$ steps at each call, thus contributing $O(n^3)$ altogether. Tarjan's algorithm and its modification both run in $O(|A| + |S|)$ which is bounded by $O(n^2)$, and moreover the arcs from different recursive steps are also different, so the overall contribution of Line 11 is $O(n^2)$.

Let us now deal with the more complex Lines 16 and 18. Let r be the number of recursive calls that are made to HubRec, and at the j -th call let T_j denote the vertices otherwise named T . Since the $(j+1)$ -th recursive call does not involve vertices in T_j , we obtain $\sum_{j=1}^r |T_j| \leq n$. The loop at Line 18 is taken at most $n^2 |T_j|$ times during the j -th call, which yields an overall contribution of $O(n^3)$. Likewise, since a basic shortest-path algorithms terminates within $O(n^2)$ steps and since it is called $|T_j|$ times during the j -th recursive call, Line 16's overall contribution is $O(n^3)$. \square

Proof [Proposition 40] Up to focusing let p satisfy Assumption 11. Let y be in the set of the transient states T , so there exist $x \notin T$ and $\gamma \in \Gamma_T(y, x)$ in the essential graph. By Assumption 11 this implies

$$0 < \liminf_{\epsilon \rightarrow 0} p_\epsilon(\gamma) \leq \liminf_{\epsilon \rightarrow 0} \mathbb{P}_\epsilon^y(\tau_x^+ < \tau_y^+).$$

On the other hand, let E be the essential class of x . Lemma 15 implies that

$$\begin{aligned}\mathbb{P}^x(\tau_y^+ < \tau_x^+) &\leq 1 - \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}^+} = x) \\ &= \sum_{z \notin E} \mathbb{P}^x(X_{\tau_{S \setminus E \cup \{x\}}^+} = z) \cong \sum_{z \notin E} \max_{t \in E} p(t, z)\end{aligned}$$

Since E is an essential class, $\sum_{z \notin E} \max_{t \in E} p_\epsilon(t, z) \xrightarrow{\epsilon \rightarrow 0} 0$, thus by Lemma 6

$$\tilde{\mu}_\epsilon(y) \leq \frac{\tilde{\mu}_\epsilon(y)}{\mu_\epsilon(x)} = \frac{\mathbb{P}_\epsilon^x(\tau_y^+ < \tau_x^+)}{\mathbb{P}_\epsilon^y(\tau_x^+ < \tau_y^+)} \xrightarrow{\epsilon \rightarrow 0} 0.$$

□

Proof [Corollary 41] In the procedure underlying Theorem 27, only the states that are transient at some point during the run are deleted. By Proposition 40 these are exactly the fully vanishing states. □

Proof [Lemma 42] For all $x \in S$ let $q_\epsilon^x := \sum_{T \in \mathcal{T}_x} \prod_{(z,t) \in T} p_\epsilon(z, t)$ and let $q := (q_\epsilon^y)_{y \in S, \epsilon \in I}$. By irreducibility $q_\epsilon^z > 0$ for all $z \in S$ and $\epsilon \in I$, so let $\mu_\epsilon^x := \frac{q_\epsilon^x}{\sum_{y \in S} q_\epsilon^y}$ for all $z \in S$ and $\epsilon \in I$. Let us assume that $\beta^y \precsim \beta^x$ for all $y \in S$, so by finiteness of S there exist positive c and ϵ_0 such that $\beta_\epsilon^y \leq c \cdot \beta_\epsilon^x$ for all $y \in S$ and $\epsilon < \epsilon_0$. For all $y \in S$ and $\epsilon < \epsilon_0$ we have

$$q_\epsilon^x \geq \beta_\epsilon^x \geq \frac{\beta_\epsilon^y}{c} \geq \frac{1}{c \cdot |\mathcal{T}_y|} \cdot \sum_{T \in \mathcal{T}_y} \prod_{(z,t) \in T} p_\epsilon(z, t) = \frac{q_\epsilon^y}{c \cdot |\mathcal{T}_y|}$$

Note that $|\mathcal{T}_y| \leq 2^{|S|^2}$ since a spanning tree of a graph is a subset of its arcs, so

$$\mu_\epsilon^x = \frac{q_\epsilon^x}{\sum_{y \in S} q_\epsilon^y} \geq \frac{1}{c \cdot \sum_{y \in S} |\mathcal{T}_y|} \geq \frac{1}{c \cdot |S| \cdot 2^{|S|^2}}$$

which ensures that $\liminf_{\epsilon \rightarrow 0} \mu_\epsilon^x > 0$. By the Markov chain tree theorem $\mu \cdot p = \mu$, so x is a stable state.

Conversely, let us assume that $\neg(\beta^y \precsim \beta^x)$ for some $y \in S$, so for all $c, \epsilon > 0$ there exists a positive $\eta < \epsilon$ such that $c \cdot \beta_\eta^x < \beta_\eta^y$. Let $c, \epsilon > 0$ and let a positive $\eta < \epsilon$ be such that $c \cdot 2^{|S|^2} \cdot \beta_\eta^x < \beta_\eta^y$, so $c \cdot \mu_\eta^x < \mu_\eta^y$. Since $\mu \leq 1$, it shows that $\liminf_{\epsilon \rightarrow 0} \mu_\epsilon^x = 0$. □

Proof [Observation 43] let G be the graph with arc (x, y) if $p(x, y) > 0$. Let $E'_1, \dots, E'_{k'}$ be the sink (aka bottom) strongly connected components of G , so a state is stable for p iff it is stable for one of the $p|_{E'_i \times E'_i}$. Since the $p|_{E'_i \times E'_i}$ are irreducible perturbations, Lemma 42 can be applied, and by Assumption 5 the weights of the spanning trees are totally preordered, so there are stable states. □

Proof [Observation 44] For all $x, y \in S$ let I_{xy} be the support of $p(x, y) : I \rightarrow [0, 1]$. By Assumption 5 the I_{xy} are totally ordered by inclusion. Among these sets let $0 \subsetneq I_1 \subsetneq \dots \subsetneq I_l \subsetneq I$ be the non-trivial subsets of I . Up to focusing on a smaller neighborhood of 0 inside I , let us assume that 0 is a limit point of I_1 , all the $I_{i+1} \setminus I_i$, and $I \setminus I_l$. By Lemma 47.8 a state is stable for p iff it is stable for $p|_{I_1}$, all the $p|_{I_{i+1} \setminus I_i}$, and $p|_{I \setminus I_l}$. These restrictions all satisfy the positivity assumption of Observation 43, whose underlying algorithm computes the stable states in $O(n^3)$. Since there are at most n^2 restrictions, stability is decidable in $O(n^5)$. □

Proof [Proposition 45] By induction on $n := |\{p(x, y) \mid x \neq y \wedge p(x, y) \neq 0 \wedge \neg(0 < p(x, y))\}|$. If $n = 0$, let G be the graph with arc (x, y) if $p(x, y) > 0$. Let $E'_1, \dots, E'_{k'}$ be the sink SCCs of G , so a state is stable for p iff it is stable for one of

the $p|_{E'_i \times E'_i}$. By decidability of \precsim and since the $p|_{E'_i \times E'_i}$ are irreducible perturbations, Lemma 42 allows us to compute their stable states.

If $n > 0$ let $p(x, y)$ be a non-zero function with zeros, and let J be its support. If 0 is not a limit point of $J(I \setminus J)$, the stable states of p are the stable states of $p|_{I \setminus J}(p_J)$, which are computable by induction hypothesis. If 0 is a limit point of both J and $I \setminus J$, by Lemma 47.8 the stable states wrt p are the states that are stable wrt both p_J and $p|_{I \setminus J}$, and we can use the induction hypothesis for both. □

Proof [Observation 46] By induction. More specifically, let us prove that these roots are preserved and reflected by outgoing scaling, essential collapse, and transient deletion.

- Since the outgoing scaling divides all the coefficients by the same scale $f \in F$, the weights of the spanning trees are all divided by $f^{|S|-1}$, and the order between them is preserved.
- Let E be a (sink) SCC of the essential graph of P , and let $x, y \in E$. It is easy to see that a spanning tree rooted at x can be modified (only within E) into a spanning tree rooted at y that has the same weight. Since the arcs in E do not contribute to the weight, the essential collapse is safe.
- Let the sink SCCs of P be singletons, and let $\{y\}$ not be one of those, so there exists a path from y to a sink SCC $\{x\}$. Let T be a spanning tree rooted at y . Following T , let x' be the successor of x , so the weight of (x, x') is less than 1. Let us modify T into T' by letting y lead to the new root x by a path of weight 1. The weight of T' is greater than that of T by at least the weight of (x, x') . This shows that only essential vertices may be the roots of spanning trees of maximum weights. Moreover, let T be a spanning tree of maximum weight, and let x and y be essential vertices such that following T from x leads to y without visiting any other essential vertex. Then this path between x and y must have maximal weight among all paths from x to y that avoid other essential vertices. So the weight of maximal spanning trees after transient deletion correspond to the weight before deletion.

□